

Fault-Tolerant Multi-Agent Optimization– Part III *

Lili Su

Nitin Vaidya

Department of Electrical and Computer Engineering, and
Coordinated Science Laboratory
University of Illinois at Urbana-Champaign
Email: {lilisu3, nhv}@illinois.edu

Technical Report

September 8, 2015

Abstract. We study fault-tolerant distributed optimization of a sum of convex (cost) functions with real-valued scalar input/output in the presence of crash faults or Byzantine faults. In particular, the goal is to optimize a global cost function $\frac{1}{n} \sum_{i \in \mathcal{V}} h_i(x)$, where $\mathcal{V} = \{1, \dots, n\}$ is the collection of agents, and $h_i(x)$ is agent i 's local cost function, which is initially known only to agent i . This problem finds its applications in the domain of fault-tolerant large scale distributed machine learning, where data are generated at different locations and some data may be lost during processing or be tampered by malicious local data managers. The global cost function $\frac{1}{n} \sum_{i \in \mathcal{V}} h_i(x)$ captures the requirement that, in distributed machine learning, the system tries to take full advantage of all the data generated at different locations. Since the above global cost function cannot be optimized exactly in presence of crash faults or Byzantine faults, we define two weaker versions of the problem for crash faults and Byzantine faults, respectively.

When some agents may crash, the local functions/data stored at these agents may not always available to the system. In this scenario, the goal for the weaker problem is to generate an output that is an optimum of a function formed as

$$C \left(\sum_{i \in \mathcal{N}} h_i(x) + \sum_{i \in \mathcal{F}} \alpha_i h_i(x) \right),$$

where \mathcal{N} is the set of non-faulty agents, \mathcal{F} is the set of faulty agents (crashed agents), $0 \leq \alpha_i \leq 1$ for each $i \in \mathcal{F}$ and C is a normalization constant such that $C(|\mathcal{N}| + \sum_{i \in \mathcal{F}} \alpha_i) = 1$. We present an iterative algorithm in which each agent only needs to perform local computation, and send one message per iteration.

When some agents may be Byzantine, the system cannot take full advantage of the data kept by non-faulty agents. The goal for the associated weaker problem is to generate an output that is an optimum of a function formed as

$$\sum_{i \in \mathcal{N}} \alpha_i h_i(x),$$

such that $\alpha_i \geq 0$ for each $i \in \mathcal{N}$ and $\sum_{i \in \mathcal{N}} \alpha_i = 1$. We present an iterative algorithm, where only local computation is needed and only one message per agent is sent in each iteration, that ensures that at least $|\mathcal{N}| - f$ agents have weights (α_i 's) that are lower bounded by $\frac{1}{2(|\mathcal{N}| - f)}$.

The obtained results can be generalized to asynchronous systems as well.

* This research is supported in part by National Science Foundation awards NSF 1329681 and 1421918. Any opinions, findings, and conclusions or recommendations expressed here are those of the authors and do not necessarily reflect the views of the funding agencies or the U.S. government.

1 System Model and Problem Formulation

The system under consideration is synchronous, and consists of n agents connected by a complete communication network. Our results can be generalized to asynchronous system. We postpone the discussion of this generalization to the end of this report. The set of agents is $\mathcal{V} = \{1, \dots, n\}$. We assume that $n > 3f$ for reasons that will be clearer soon. We say that a function $h : \mathbb{R} \rightarrow \mathbb{R}$ is *admissible* if (i) $h(\cdot)$ is convex, and continuously differentiable, (ii) the set $\arg \min_{x \in \mathbb{R}} h(x)$ containing the optima of $h(\cdot)$ is non-empty and compact (i.e., bounded and closed), (iii) the magnitude of the gradient is bounded by L , i.e., $|h'(x)| \leq L, \forall x \in \mathbb{R}$, and the derivative $h'(\cdot)$ is L -Lipschitz continuous. Each agent $i \in \mathcal{V}$ is initially provided with an *admissible* local cost function $h_i : \mathbb{R} \rightarrow \mathbb{R}$. Ideally, the system goal is to optimize the *average* of all the local functions, and have all the agents to reach agreement on the optimum x . In particular, each agent should output an *identical* value $\tilde{x} \in \mathbb{R}$ that minimizes

$$\frac{1}{n} \sum_{i \in \mathcal{V}} h_i(x). \quad (1)$$

This problem finds its applications in the domain of large scale distributed machine learning, where data are generated at different locations and the data center at each location is not allowed to transmit all the locally collected data to other centers either due to transmission capacity constraint or due to privacy issue.

This problem is well-studied in the scenario where each agent is reliable throughout any execution of an algorithm [7,15,21]. In this work, we consider the fault-tolerant version of this problem. In particular, we consider the setting where up to f of the n agents may crash or be Byzantine faulty. Let \mathcal{F} denote the set of faulty agents, and let $\mathcal{N} = \mathcal{V} - \mathcal{F}$ denote the set of non-faulty agents. For each $t \geq 0$, let $\mathcal{N}[t]$ be the collection of agents that have not been crashed till the end of iteration t , with $\mathcal{N}[0] = \mathcal{V}$. Note that $\mathcal{N}[t+1] \subseteq \mathcal{N}[t]$ for $t \geq 0$, and that $\lim_{t \rightarrow \infty} \mathcal{N}[t] = \mathcal{N}$. The set \mathcal{F} of faulty agents may be chosen by an adversary arbitrarily. Let $|\mathcal{F}| = \phi$. Note that $\phi \leq f$ and $|\mathcal{N}| \geq n - f$. The presence of crashed or Byzantine faulty agents makes it impossible to design an algorithm that solves (1) for all admissible local cost functions (this is shown formally in Part I of this report [19]). Therefore, for crash fault and Byzantine fault, respectively, we study two weaker versions of the problem, namely, Problem 1 and Problem 2 in Figure 1. Problem 1 is proposed in this report, and Problem 2 is first introduced in Part I of this report [19].

When some agents may crash, the local functions/data stored at these agents may only be partially visible or even invisible to the system – as agent i may crash at any time during an execution. Problem 1 requires that the output \tilde{x} be an optimum of a function formed as

$$C \left(\sum_{i \in \mathcal{N}} h_i(x) + \sum_{i \in \mathcal{F}} \alpha_i h_i(x) \right),$$

where \mathcal{N} is the set of non-faulty agents, \mathcal{F} is the set of faulty agents (crashed agents), $0 \leq \alpha_i \leq 1$ for each $i \in \mathcal{F}$ and C is a normalization constant such that $C (|\mathcal{N}| + \sum_{i \in \mathcal{F}} \alpha_i) = 1$.

When some agents may be Byzantine, the system cannot take full advantage of the data kept by non-faulty agents. In addition, among the non-faulty agents, the system may put more weights to some agents than the others. Then, the desired goal is to *maximize* the number of weights (α_i 's) that are bounded away from zero. With this in mind, Problem 2 in Figure 1 is introduced in Part I of this report [19]. In Problem 2, note that $\mathbf{1}\{\alpha_i > \beta\}$ is an indicator function that outputs 1 if $\alpha_i > \beta$, and 0 otherwise. Essentially, Problem 2 requires that at least γ weights must exceed a

threshold β , where $\beta \geq 0$. Thus, β, γ are parameters of Problem 2, capturing how the data collected by non-faulty agents are utilized by the system.

Problem 1	Problem 2 with parameters $\beta, \gamma, \beta \geq 0$
$\tilde{x} \in \arg \min_{x \in \mathbb{R}} C \left(\sum_{i \in \mathcal{N}} h_i(x) + \sum_{i \in \mathcal{F}} \alpha_i h_i(x) \right)$ <p>such that</p> $\forall i \in \mathcal{F}, 0 \leq \alpha_i \leq 1 \text{ and}$ $C \left(\mathcal{N} + \sum_{i \in \mathcal{F}} \alpha_i \right) = 1$	$\tilde{x} \in \arg \min_{x \in \mathbb{R}} \sum_{i \in \mathcal{N}} \alpha_i h_i(x)$ <p>such that</p> $\forall i \in \mathcal{N}, \alpha_i \geq 0,$ $\sum_{i \in \mathcal{N}} \alpha_i = 1, \text{ and}$ $\sum_{i \in \mathcal{N}} \mathbf{1}(\alpha_i > \beta) \geq \gamma$

Fig. 1: Problem formulations: All non-faulty agents must output an identical value $\tilde{x} \in \mathbb{R}$ that satisfies the constraints specified in each problem formulation.

We will say that Problem 1 or 2 is solvable if there exists an algorithm that will find a solution for the problem (satisfying all its constraints) for all admissible local cost functions, and all possible behaviors of faulty agents. Our problem formulations require that all non-faulty agents output asymptotically identical $\tilde{x} \in \mathbb{R}$, while satisfying the constraints imposed by the problem (as listed in Figure 1). Thus, the traditional fault-tolerant consensus [9] problem, which also imposes a similar *agreement* condition, is a special case of our optimization problem.¹ Therefore, the lower bound of $n > 3f$ for Byzantine consensus [9] also applies to our problem. Hence we assume that $n > 3f$.

We prove the following key results:

- (Theorem 2) We provide a simple iterative algorithm that solves Problem 1. In each iteration of this algorithm, each agent only needs to perform local computation and send one message.
- (Theorem 4) We present a simple iterative algorithm that solves Problem 2 with $\beta \leq \frac{1}{2(|\mathcal{N}| - f)}$ and $\gamma \leq |\mathcal{N}| - f$. In each iteration of this algorithm, each agent only needs to perform local computation and send one message.

In our proposed algorithms, the local estimates at all non-faulty agents are identical in the limit.

The rest of the report is organized as follows. Related work is summarized in Section 2. Two algorithms are proposed in Section 3, wherein the first algorithm solves Problem 1 with two rounds of information exchange in each iteration. In contrast, the second algorithm solves Problem 1 with one message sent per agent in each iteration. Section 4 presents a simple iterative algorithm that solves Problem 2 with $\beta = \frac{1}{2(|\mathcal{N}| - f)}$ and $\gamma = |\mathcal{N}| - f$. Similar to the second algorithm in Section 3, this proposed algorithm only requires one message sent per agent in each iteration. Section 5 discusses the generalization of the obtained results to asynchronous systems, and concludes the report.

¹ Interested readers are referred to Part I of this report [19] for formal proof.

2 Related Work

Fault-tolerant consensus [16] is a special case of the optimization problem considered in this report. There is a significant body of work on fault-tolerant consensus, including [6,5,14,8,12,23,10]. The optimization algorithms presented in this report use Byzantine consensus as a component.

Convex optimization, including distributed convex optimization, also has a long history [3]. However, we are not aware of prior work that obtains the results presented in this report except [19,20]. Primal and dual decomposition methods that led themselves naturally to a distributed paradigm are well-known [4]. There has been significant research on a variant of distributed optimization problem [7,15,21], in which the global objective $h(x)$ is a summation of n convex functions, i.e., $h(x) = \sum_{j=1}^n h_j(x)$, with function $h_j(x)$ being known to the j -th agent. The need for robustness for distributed optimization problems has received some attentions recently [7,11,24,13,19,20]. In particular, Duchi et al. [7] studied the impact of random communication link faults on the convergence of distributed variant of dual averaging algorithm. Specifically, each realizable link fault pattern considered in [7] is assumed to admit a doubly-stochastic matrix which governs the evolution dynamics of local estimates of the optimum.

We considered Byzantine fault in [19] and [20]. Both [19] and [20] considered synchronous system. [19] showed that at most $|\mathcal{N}| - f$ non-faulty functions can have non-zero weights. This observation led to the formulation of Problem 2 in Fig. 1. Six algorithms were proposed in [19]. In contrast, we also showed [20] that sufficient redundancy in the input functions (each input function is not exclusively kept by a single agent), it is possible to solve (1), where the summation is over all input functions. In addition, a simple low-complexity iterative algorithm was proposed in [20], and a tight topological condition for the existence of such iterative algorithms is identified.

In other related work, significant attempts have been made to solve the problem of distributed hypothesis testing in the presence of Byzantine attacks [11,24,13], where Byzantine sensors may transmit fictitious observations aimed at confusing the decision maker to arrive at a judgment that is in contrast with the true underlying distribution. Consensus based variant of distributed event detection, where a centralized data fusion center does not exist, is considered in [11]. In contrast, in this paper, we focus on the Byzantine attacks on the multi-agent optimization problem.

3 Mutil-Agent Optimization with Crash fault

Algorithm 1 and its correctness proof contain the key ideas and intuition of this report.

3.1 Algorithm 1: Two-Round of Information Exchange per Iteration

In Algorithm 1, each agent j maintains two variables: the local estimate x_j and the auxiliary variable s_j , with $x_j[t]$ and $s_j[t]$ representing these two variables at the end of iteration t , and $x_j[0]$ being the system input at agent j and $s_j[0] = 0$. In each iteration $t \geq 1$, there are two rounds of information exchange. In the first round, (1) agent j requests all the agents (including itself) to compute the gradients of their local functions at $x_j[t-1]$; (2) after receiving $x_j[t-1]$, a non-faulty agent i computes $h'_i(x_j[t-1])$ and sends it back to agent j ; (3) agent j collects the requested gradients and updates the auxiliary variable s_j . In the second round, all the non-faulty agents exchange their auxiliary variables $s_j[t]$'s and update their local estimate as an *average* of all received auxiliary variables.

Let $\{\lambda[t]\}_{t=0}^\infty$ be a sequence of stepsizes chosen beforehand such that $\lambda[t] \geq 0$ and $\lambda[t] \geq \lambda[t+1]$ for each $t \geq 0$, $\lim_{t \rightarrow \infty} \lambda[t] = 0$, $\sum_{t=0}^\infty \lambda[t] = \infty$ and $\sum_{t=0}^\infty \lambda^2[t] < \infty$.

Algorithm 1 for agent j at iteration t :

Step 1: Send $x_j[t-1]$ to all the agents (including agent j itself).

Step 2: Upon receiving $x_i[t-1]$ from agent i , compute $h'_j(x_i[t-1])$ – the gradient of function $h_j(\cdot)$ at $x_i[t-1]$ – and send it back to agent i .

Step 3: Let $\mathcal{R}_j^1[t-1]$ denote the set of gradients of the form $h'_i(x_j[t-1])$ received as a result of step 1 and step 2. Update s_j as

$$s_j[t] = x_j[t-1] - \frac{\lambda[t-1]}{|\mathcal{R}_j^1[t-1]|} \left(\sum_{i \in \mathcal{R}_j^1[t-1]} h'_i(x_j[t-1]) \right). \quad (2)$$

Step 4: Send $s_j[t]$ to all the agents (including agent j itself).

Step 5: Let $\mathcal{R}_j^2[t-1]$ denote the set of auxiliary variables $s_i[t]$ received as a result of step 4. Update x_j as

$$x_j[t] = \frac{1}{|\mathcal{R}_j^2[t-1]|} \sum_{i \in \mathcal{R}_j^2[t-1]} s_i[t]. \quad (3)$$

Steps 1, 2 and 3 correspond to the first round of information exchange, and step 4 corresponds to the second round of information exchange. We will show that Algorithm 1 correctly solves Problem 1. Intuitively speaking, the first round of information exchange corresponds to the standard gradient-method iterate, which drives each local estimate to a global optimum; the second round of information exchange forces all local estimates at non-faulty agents to reach consensus. Algorithm 2 will achieve a similar goal with a single round of exchange.

Recall that \mathcal{N} is the set of non-faulty agents and \mathcal{F} is the set of faulty agents that may crash at any time during an execution. In Problem 1, the system goal is to optimize

$$C \left(\sum_{i \in \mathcal{N}} h_i(x) + \sum_{i \in \mathcal{F}} \alpha_i h_i(x) \right),$$

where \mathcal{N} is the set of non-faulty agents, \mathcal{F} is the set of faulty agents (crashed agents), $0 \leq \alpha_i \leq 1$ for each $i \in \mathcal{F}$ and C is a normalization constant such that $C(|\mathcal{N}| + \sum_{i \in \mathcal{F}} \alpha_i) = 1$. Given \mathcal{N} and \mathcal{F} , the normalization constant C and the crashed agents' coefficients α_i depend on when the faulty agents crash during an execution. For given \mathcal{N} and \mathcal{F} , let \mathcal{C} be the collection of potential system objectives, formally defined as follows:

$$\mathcal{C} \triangleq \left\{ p(x) : p(x) = C \left(\sum_{i \in \mathcal{N}} h_i(x) + \sum_{i \in \mathcal{F}} \alpha_i h_i(x) \right), \right. \\ \left. \forall i \in \mathcal{F}, 0 \leq \alpha_i \leq 1, C \left(|\mathcal{N}| + \sum_{i \in \mathcal{F}} \alpha_i \right) = 1. \right\} \quad (4)$$

Each $p(x) \in \mathcal{C}$ is called a valid function. Since $\forall i \in \mathcal{F}$, $0 \leq \alpha_i \leq 1$, it holds that $\frac{1}{n} \leq C \leq \frac{1}{|\mathcal{N}|}$ for each valid function. Note that $\frac{1}{|\mathcal{N}|} \sum_{i \in \mathcal{N}} h_i(x) \in \mathcal{C}$ is a valid function. For ease of future reference, we let $\tilde{p}(x) \triangleq \frac{1}{|\mathcal{N}|} \sum_{i \in \mathcal{N}} h_i(x)$. Define $Y \triangleq \cup_{p(x) \in \mathcal{C}} \text{argmin } p(x)$. The characterization of Y is presented in the following two lemmas.

Lemma 1. *Y is a convex set.*

Lemma 2. *Y is a closed set.*

The proofs of Lemma 1 and Lemma 2 are presented in Appendix A and Appendix B, respectively.

In addition, since Y is convex, $\text{Dist}(\cdot, Y)$ is also convex.

Asymptotic Consensus under Algorithm 1 We first show that asymptotic consensus among the non-faulty agents is achieved under Algorithm 1. The following proposition is used in proving consensus.

Proposition 1. *Let $0 \leq b < 1$. Define $\ell(t) = \sum_{r=0}^{t-1} \lambda[r] b^{t-r}$. The limit of $\ell(t)$ exists and*

$$\lim_{t \rightarrow \infty} \ell(t) = 0.$$

Proposition 1 is proved in Appendix C.

Recall that for each $t \geq 0$, $\mathcal{N}[t]$ is the collection of agents that have not been crashed till the end of iteration t . Note that $\mathcal{N}[t+1] \subseteq \mathcal{N}[t]$ for $t \geq 0$, and that $\lim_{t \rightarrow \infty} \mathcal{N}[t] = \mathcal{N}$. Denote $M(t) = \max_{i \in \mathcal{N}[t]} x_i[t]$ and $m(t) = \min_{i \in \mathcal{N}[t]} x_i[t]$.

Lemma 3. *Under Algorithm 1, the sequence $\{M[t] - m[t]\}_{t=0}^{\infty}$ converges and*

$$\lim_{t \rightarrow \infty} (M[t] - m[t]) = 0.$$

Proof. Let $i, j \in \mathcal{N}[t]$ such that $x_i[t] = M[t]$, and $x_j[t] = m[t]$.

$$\begin{aligned} M[t] - m[t] &= x_i[t] - x_j[t] \\ &= \frac{1}{|\mathcal{R}_i^2[t-1]|} \sum_{k \in \mathcal{R}_i^2[t-1]} s_k[t] - \frac{1}{|\mathcal{R}_j^2[t-1]|} \sum_{p \in \mathcal{R}_j^2[t-1]} s_p[t] \quad \text{by (3)} \\ &= \min \left\{ \frac{1}{|\mathcal{R}_i^2[t-1]|}, \frac{1}{|\mathcal{R}_j^2[t-1]|} \right\} \left(\sum_{k \in \mathcal{R}_i^2[t-1]} s_k[t] - \sum_{p \in \mathcal{R}_j^2[t-1]} s_p[t] \right) \\ &\quad + \left(\frac{1}{|\mathcal{R}_i^2[t-1]|} - \min \left\{ \frac{1}{|\mathcal{R}_i^2[t-1]|}, \frac{1}{|\mathcal{R}_j^2[t-1]|} \right\} \right) \sum_{k \in \mathcal{R}_i^2[t-1]} s_k[t] \\ &\quad - \left(\frac{1}{|\mathcal{R}_j^2[t-1]|} - \min \left\{ \frac{1}{|\mathcal{R}_i^2[t-1]|}, \frac{1}{|\mathcal{R}_j^2[t-1]|} \right\} \right) \sum_{p \in \mathcal{R}_j^2[t-1]} s_p[t]. \quad (5) \end{aligned}$$

Assume that $|\mathcal{R}_i^2[t-1]| \geq |\mathcal{R}_j^2[t-1]|$. The case that $|\mathcal{R}_i^2[t-1]| < |\mathcal{R}_j^2[t-1]|$ can be shown similarly.

We can simplify (5) as follows.

$$M[t] - m[t] = \frac{1}{|\mathcal{R}_i^2[t-1]|} \left(\sum_{k \in \mathcal{R}_i^2[t-1]} s_k[t] - \sum_{p \in \mathcal{R}_j^2[t-1]} s_p[t] \right) - \left(\frac{1}{|\mathcal{R}_j^2[t-1]|} - \frac{1}{|\mathcal{R}_i^2[t-1]|} \right) \sum_{p \in \mathcal{R}_j^2[t-1]} s_p[t]. \quad (6)$$

For each k , we get

$$\begin{aligned} s_k[t] &= x_k[t-1] - \frac{\lambda[t-1]}{|\mathcal{R}_k^1[t-1]|} \left(\sum_{i \in \mathcal{R}_k^1[t-1]} h'_i(x_k[t-1]) \right) \\ &\leq x_k[t-1] + \frac{\lambda[t-1]}{|\mathcal{R}_k^1[t-1]|} \left(\sum_{i \in \mathcal{R}_k^1[t-1]} L \right) \quad \text{since } |h'_i(x)| \leq L, \forall x \in \mathbb{R}, \forall i \in \mathcal{V} \\ &= x_k[t-1] + \frac{\lambda[t-1]}{|\mathcal{R}_k^1[t-1]|} |\mathcal{R}_k^1[t-1]| L \\ &= x_k[t-1] + \lambda[t-1]L \leq M[t-1] + \lambda[t-1]L. \end{aligned} \quad (7)$$

Similarly, for each $k \in \mathcal{N}[t]$, it holds that

$$s_k[t] \geq m[t-1] - \lambda[t-1]L. \quad (8)$$

We bound the two terms in the right hand side of (6) separately. For the first term of (6), we get

$$\begin{aligned} &\frac{1}{|\mathcal{R}_i^2[t-1]|} \left(\sum_{k \in \mathcal{R}_i^2[t-1]} s_k[t] - \sum_{p \in \mathcal{R}_j^2[t-1]} s_p[t] \right) \\ &= \frac{1}{|\mathcal{R}_i^2[t-1]|} \left(\sum_{k \in \mathcal{R}_i^2[t-1] \cap \mathcal{R}_j^2[t-1]} s_k[t] + \sum_{k \in \mathcal{R}_i^2[t-1] - \mathcal{R}_j^2[t-1]} s_k[t] - \sum_{p \in \mathcal{R}_j^2[t-1] \cap \mathcal{R}_i^2[t-1]} s_p[t] - \sum_{p \in \mathcal{R}_j^2[t-1] - \mathcal{R}_i^2[t-1]} s_p[t] \right) \\ &= \frac{1}{|\mathcal{R}_i^2[t-1]|} \left(\sum_{k \in \mathcal{R}_i^2[t-1] - \mathcal{R}_j^2[t-1]} s_k[t] - \sum_{p \in \mathcal{R}_j^2[t-1] - \mathcal{R}_i^2[t-1]} s_p[t] \right) \\ &\stackrel{(a)}{\leq} \frac{1}{|\mathcal{R}_i^2[t-1]|} \left(\sum_{k \in \mathcal{R}_i^2[t-1] - \mathcal{R}_j^2[t-1]} (M[t-1] + \lambda[t-1]L) - \sum_{p \in \mathcal{R}_j^2[t-1] - \mathcal{R}_i^2[t-1]} (m[t-1] - \lambda[t-1]L) \right) \\ &= \frac{|\mathcal{R}_i^2[t-1] - \mathcal{R}_j^2[t-1]|}{|\mathcal{R}_i^2[t-1]|} (M[t-1] + \lambda[t-1]L) - \frac{|\mathcal{R}_j^2[t-1] - \mathcal{R}_i^2[t-1]|}{|\mathcal{R}_i^2[t-1]|} (m[t-1] - \lambda[t-1]L). \end{aligned} \quad (9)$$

Inequality (a) holds due to (7) and (8).

For the second term of (6), we get

$$\begin{aligned}
-\left(\frac{1}{|\mathcal{R}_j^2[t-1]|} - \frac{1}{|\mathcal{R}_i^2[t-1]|}\right) \sum_{p \in \mathcal{R}_j^2[t-1]} s_p[t] &= -\frac{|\mathcal{R}_i^2[t-1]| - |\mathcal{R}_j^2[t-1]|}{|\mathcal{R}_j^2[t-1]| |\mathcal{R}_i^2[t-1]|} \sum_{p \in \mathcal{R}_j^2[t-1]} s_p[t] \\
&\leq -\frac{|\mathcal{R}_i^2[t-1]| - |\mathcal{R}_j^2[t-1]|}{|\mathcal{R}_j^2[t-1]| |\mathcal{R}_i^2[t-1]|} \sum_{p \in \mathcal{R}_j^2[t-1]} (m[t-1] - \lambda[t-1]L) \\
&= -\frac{|\mathcal{R}_i^2[t-1]| - |\mathcal{R}_j^2[t-1]|}{|\mathcal{R}_i^2[t-1]|} (m[t-1] - \lambda[t-1]L).
\end{aligned} \tag{10}$$

By (9) and (10), (6) can be further bounded as

$$\begin{aligned}
M[t] - m[t] &= \frac{1}{|\mathcal{R}_i^2[t-1]|} \left(\sum_{k \in \mathcal{R}_i^2[t-1]} s_k[t] - \sum_{p \in \mathcal{R}_j^2[t-1]} s_p[t] \right) - \left(\frac{1}{|\mathcal{R}_j^2[t-1]|} - \frac{1}{|\mathcal{R}_i^2[t-1]|} \right) \sum_{p \in \mathcal{R}_j^2[t-1]} s_p[t] \quad \text{by (6)} \\
&\leq \frac{|\mathcal{R}_i^2[t-1] - \mathcal{R}_j^2[t-1]|}{|\mathcal{R}_i^2[t-1]|} (M[t-1] + \lambda[t-1]L) - \frac{|\mathcal{R}_j^2[t-1] - \mathcal{R}_i^2[t-1]|}{|\mathcal{R}_i^2[t-1]|} (m[t-1] - \lambda[t-1]L) \\
&\quad - \frac{|\mathcal{R}_i^2[t-1]| - |\mathcal{R}_j^2[t-1]|}{|\mathcal{R}_i^2[t-1]|} (m[t-1] - \lambda[t-1]L) \quad \text{by (9) and (10)} \\
&\stackrel{(a)}{=} \frac{|\mathcal{R}_i^2[t-1] - \mathcal{R}_j^2[t-1]|}{|\mathcal{R}_i^2[t-1]|} (M[t-1] + \lambda[t-1]L) - \frac{|\mathcal{R}_i^2[t-1] - \mathcal{R}_j^2[t-1]|}{|\mathcal{R}_i^2[t-1]|} (m[t-1] - \lambda[t-1]L) \\
&= \frac{|\mathcal{R}_i^2[t-1] - \mathcal{R}_j^2[t-1]|}{|\mathcal{R}_i^2[t-1]|} (M[t-1] - m[t-1] + 2\lambda[t-1]L) \\
&\stackrel{(b)}{\leq} \frac{f}{n-f} (M[t-1] - m[t-1] + 2\lambda[t-1]L) \\
&\leq \left(\frac{f}{n-f} \right)^t (M[0] - m[0]) + 2L \left(\sum_{r=0}^{t-1} \left(\frac{f}{n-f} \right)^{t-r} \lambda[r] \right).
\end{aligned} \tag{11}$$

Equality (a) is true because that

$$\begin{aligned}
&|\mathcal{R}_j^2[t-1] - \mathcal{R}_i^2[t-1]| + |\mathcal{R}_i^2[t-1]| - |\mathcal{R}_j^2[t-1]| \\
&= |\mathcal{R}_j^2[t-1]| - |\mathcal{R}_j^2[t-1] \cap \mathcal{R}_i^2[t-1]| + |\mathcal{R}_i^2[t-1]| - |\mathcal{R}_j^2[t-1]| \\
&= |\mathcal{R}_i^2[t-1]| - |\mathcal{R}_j^2[t-1] \cap \mathcal{R}_i^2[t-1]| \\
&= |\mathcal{R}_i^2[t-1] - \mathcal{R}_j^2[t-1]|.
\end{aligned}$$

Since $|\mathcal{F}| \leq f$, it holds that $|\mathcal{R}_i^2[t-1] - \mathcal{R}_j^2[t-1]| \leq f$ and that $|\mathcal{R}_i^2[t-1]| \geq n - f$. Thus,

$$\frac{|\mathcal{R}_i^2[t-1] - \mathcal{R}_j^2[t-1]|}{|\mathcal{R}_i^2[t-1]|} \leq \frac{f}{n-f},$$

and inequality (b) holds.

It follows from Proposition 1 that

$$\lim_{t \rightarrow \infty} 2L \left(\sum_{r=0}^{t-1} \left(\frac{f}{n-f} \right)^{t-r} \lambda[r] \right) = 0,$$

where $b = \frac{f}{n-f}$. Thus, taking limit sup on both sides of (11), we get

$$\limsup_{t \rightarrow \infty} (M[t] - m[t]) \leq \lim_{t \rightarrow \infty} \left(\left(\frac{f}{n-f} \right)^t (M[0] - m[0]) \right) + 2L \lim_{t \rightarrow \infty} \left(\sum_{r=0}^{t-1} \left(\frac{f}{n-f} \right)^{t-r} \lambda[r] \right) = 0.$$

On the other hand, by definition of $(M[t] - m[t])$, for each $t \geq 0$ we get $(M[t] - m[t]) \geq 0$. Thus

$$\liminf_{t \rightarrow \infty} (M[t] - m[t]) \geq 0.$$

Then, we obtain

$$\limsup_{t \rightarrow \infty} (M[t] - m[t]) \leq 0 \leq \liminf_{t \rightarrow \infty} (M[t] - m[t]).$$

Thus, the limit of $(M[t] - m[t])$ exists and

$$\lim_{t \rightarrow \infty} (M[t] - m[t]) = 0.$$

□

Recall that $M[t] = \max_{i \in \mathcal{N}[t]} x_i[t]$ and $m[t] = \min_{i \in \mathcal{N}[t]} x_i[t]$. Lemma 3 implies that asymptotic consensus is achieved under Algorithm 1. The following lemma is used in the correctness proof of Algorithm 1.

Lemma 4. *Under Algorithm 1, the following holds.*

$$\sum_{t=0}^{\infty} \lambda[t] (M[t] - m[t]) < \infty.$$

Proof. Since

$$\sum_{t=0}^{\infty} \lambda[t] (M[t] - m[t]) = \lambda[0] (M[0] - m[0]) + \sum_{t=1}^{\infty} \lambda[t] (M[t] - m[t]),$$

and $\lambda[0] (M[0] - m[0]) < \infty$, to show Lemma 4, it is enough to show that

$$\sum_{t=1}^{\infty} \lambda[t] (M[t] - m[t]) < \infty.$$

$$\begin{aligned}
\sum_{t=1}^{\infty} \lambda[t] (M[t] - m[t]) &\leq \sum_{t=1}^{\infty} \lambda[t] \left(\left(\frac{f}{n-f} \right)^t (M[0] - m[0]) + 2L \left(\sum_{r=0}^{t-1} \left(\frac{f}{n-f} \right)^{t-r} \lambda[r] \right) \right) \quad \text{by (11)} \\
&= (M[0] - m[0]) \sum_{t=1}^{\infty} \lambda[t] \left(\frac{f}{n-f} \right)^t + 2L \sum_{t=1}^{\infty} \sum_{r=0}^{t-1} \left(\left(\frac{f}{n-f} \right)^{t-r} \lambda[r] \lambda[t] \right) \\
&\stackrel{(a)}{\leq} (M[0] - m[0]) \sum_{t=1}^{\infty} \lambda[t] \left(\frac{f}{n-f} \right)^t + L \sum_{t=1}^{\infty} \sum_{r=0}^{t-1} \left(\left(\frac{f}{n-f} \right)^{t-r} (\lambda^2[r] + \lambda^2[t]) \right) \\
&= (M[0] - m[0]) \sum_{t=1}^{\infty} \lambda[t] \left(\frac{f}{n-f} \right)^t + L \sum_{t=1}^{\infty} \lambda^2[t] \sum_{r=0}^{t-1} \left(\frac{f}{n-f} \right)^{t-r} \\
&\quad + L \sum_{t=1}^{\infty} \sum_{r=0}^{t-1} \left(\left(\frac{f}{n-f} \right)^{t-r} \lambda^2[r] \right) \tag{12}
\end{aligned}$$

Inequality (a) holds because $\lambda[t]\lambda[r] \leq \frac{\lambda^2[t] + \lambda^2[r]}{2}$. We bound the three terms in the RHS of (12) separately.

The first term of (12): Since $\lambda[t] \leq \lambda[0]$ for each $t \geq 1$, we have

$$\begin{aligned}
(M[0] - m[0]) \sum_{t=1}^{\infty} \lambda[t] \left(\frac{f}{n-f} \right)^t &\leq (M[0] - m[0]) \lambda[0] \sum_{t=1}^{\infty} \left(\frac{f}{n-f} \right)^t \\
&\leq (M[0] - m[0]) \lambda[0] \frac{1}{1 - \frac{f}{n-f}} \\
&= (M[0] - m[0]) \lambda[0] \frac{n-f}{n-2f} < \infty. \tag{13}
\end{aligned}$$

The second term of (12):

$$\begin{aligned}
L \sum_{t=1}^{\infty} \lambda^2[t] \sum_{r=0}^{t-1} \left(\frac{f}{n-f} \right)^{t-r} &= L \sum_{t=1}^{\infty} \lambda^2[t] \sum_{r=1}^t \left(\frac{f}{n-f} \right)^r \\
&\leq L \sum_{t=1}^{\infty} \lambda^2[t] \sum_{r=0}^{\infty} \left(\frac{f}{n-f} \right)^r \\
&= L \sum_{t=1}^{\infty} \lambda^2[t] \frac{1}{1 - \frac{f}{n-f}} \\
&= \frac{n-f}{n-2f} L \sum_{t=1}^{\infty} \lambda^2[t] \\
&< \infty \tag{14}
\end{aligned}$$

The last inequality follows from the fact that $\sum_{t=1}^{\infty} \lambda^2[t] \leq \sum_{t=0}^{\infty} \lambda^2[t] < \infty$.

The third term of (12): For any fixed T , we get

$$\begin{aligned} L \sum_{t=1}^T \sum_{r=0}^{t-1} \left(\left(\frac{f}{n-f} \right)^{t-r} \lambda^2[r] \right) &= L \sum_{r=0}^{T-1} \lambda^2[r] \sum_{t=1}^T \left(\frac{f}{n-f} \right)^t \\ &\leq L \sum_{r=0}^{T-1} \lambda^2[r] \sum_{t=0}^{\infty} \left(\frac{f}{n-f} \right)^t \\ &= \frac{n-f}{n-2f} L \sum_{r=0}^{T-1} \lambda^2[r]. \end{aligned}$$

Let $T \rightarrow \infty$, we get

$$L \sum_{t=1}^{\infty} \sum_{r=0}^{t-1} \left(\left(\frac{f}{n-f} \right)^{t-r} \lambda^2[r] \right) = \frac{n-f}{n-2f} L \sum_{r=0}^{\infty} \lambda^2[r] < \infty. \quad (15)$$

We get

$$\begin{aligned} \sum_{t=1}^{\infty} \lambda[t] (M[t] - m[t]) &\leq (M[0] - m[0]) \sum_{t=1}^{\infty} \lambda[t] \left(\frac{f}{n-f} \right)^t + L \sum_{t=1}^{\infty} \lambda^2[t] \sum_{r=0}^{t-1} \left(\frac{f}{n-f} \right)^{t-r} \\ &\quad + L \sum_{t=1}^{\infty} \sum_{r=0}^{t-1} \left(\left(\frac{f}{n-f} \right)^{t-r} \lambda^2[r] \right) \quad \text{by (12)} \\ &< \infty + \infty + \infty = \infty \quad \text{by (13), (14) and (15)} \end{aligned}$$

proving the lemma. □

By Lemma 4, we know there exists some constant C_1 such that for any constant $t \geq 0$

$$\sum_{\tau=t}^{\infty} \lambda[\tau] L (M[\tau] - m[\tau]) \leq \sum_{\tau=0}^{\infty} \lambda[\tau] L (M[\tau] - m[\tau]) \leq C_1. \quad (16)$$

The following corollary is an immediate consequence of Lemma 4.

Corollary 1. *Under Algorithm 1,*

$$\lim_{t \rightarrow \infty} \lambda[t] (M[t] - m[t]) = 0, \quad (17)$$

and

$$\lim_{t \rightarrow \infty} \sum_{\tau=t}^{\infty} \lambda[\tau] (M[\tau] - m[\tau]) = 0. \quad (18)$$

Proof. By Lemma 4, (17) holds trivially. Now we prove (18).

Let $F = \sum_{\tau=0}^{\infty} \lambda[\tau] (M[\tau] - m[\tau])$, and let $\{F_t\}_{t=0}^{\infty}$ be a sequence such that for each t ,

$$F_t = \sum_{\tau=0}^{t-1} \lambda[\tau] (M[\tau] - m[\tau]).$$

Since $M[\tau] - m[\tau] \geq 0$ for each $\tau \geq 0$, by construction, it holds that $F_t \leq F_{t+1}$ and that $F_t \leq F$ for each $t \geq 0$. Thus, by MCT, we know that

$$\lim_{t \rightarrow \infty} F_t = F.$$

Now, let

$$R_t \triangleq F - F_t = \sum_{\tau=0}^{\infty} \lambda[\tau](M[\tau] - m[\tau]) - \sum_{\tau=0}^{t-1} \lambda[\tau](M[\tau] - m[\tau]) = \left(\sum_{\tau=t}^{\infty} \lambda[\tau](M[\tau] - m[\tau]) \right).$$

By Lemma 4, we know that $F < \infty$. Thus the sequence R_t is well-defined. In addition, since the sequence F_t converges, then the sequence R_t also converges. So, we get

$$\lim_{t \rightarrow \infty} \left(\sum_{\tau=t}^{\infty} \lambda[\tau](M[\tau] - m[\tau]) \right) = \lim_{t \rightarrow \infty} R_t = \lim_{t \rightarrow \infty} (F - F_t) = F - \lim_{t \rightarrow \infty} F_t = F - F = 0,$$

proving (18). □

Optimality of Algorithm 1

Definition 1. Given a sequence $\{x[t]\}_{t=0}^{\infty}$, a sequence of gradients $\{g[t]\}_{t=0}^{\infty}$, and a set of stepsizes $\{\lambda[t]\}_{t=0}^{\infty}$ we say $x[t]$ is a resilient point with respect to gradient $g[t]$ if one of the following items is true:

- * $x[t] \in Y$ and $(x[t] - \lambda[t]g[t]) \notin Y$,
- * $x[t] > \max Y$ and $(x[t] - \lambda[t]g[t]) < \min Y$,
- * $x[t] < \min Y$ and $(x[t] - \lambda[t]g[t]) > \max Y$.

Since by Lemma 2, we know that set Y is closed. Thus $\max Y$ and $\min Y$ exist, and Definition 1 is well-defined over set Y .

Let $\{z[t]\}_{t=0}^{\infty}$ be a sequence of estimates such that

$$z[t] = x_{j_t}[t], \text{ where } j_t \in \operatorname{argmax}_{j \in \mathcal{N}[t]} \operatorname{Dist}(x_j[t], Y). \quad (19)$$

From the definition, there is a sequence of agents $\{j_t\}_{t=0}^{\infty}$ associated with the sequence $\{z[t]\}_{t=0}^{\infty}$.

Lemma 5. If there exists $c > 0$ such that

$$\lim_{t \rightarrow \infty} \operatorname{Dist}(z[t], Y) = c,$$

then at least one of the following two statements is true.

(A.1) There exists a subsequence $\{z[t_k]\}_{k=0}^{\infty}$ such that $z[t_k] < \min Y$ for all $k \geq 0$.

(A.2) There exists a subsequence $\{z[t'_k]\}_{k=0}^{\infty}$ such that $z[t'_k] > \max Y$ for all $k \geq 0$.

In addition, at least one of $(\min Y - c)$ or $(\max Y + c)$ is an accumulation point of $\{z[t]\}_{t=0}^{\infty}$.

Proof. Since $\lim_{t \rightarrow \infty} \operatorname{Dist}(z[t], Y) = c > 0$, there exists m such that $z[t] \notin Y$ for each $t \geq m$. Otherwise, there exists a subsequence $\{z[t_k]\}_{k=0}^{\infty}$ such that $z[t_k] \in Y$ for each $k \geq 0$. By definition of $\operatorname{Dist}(\cdot, Y)$, we have, $\operatorname{Dist}(z[t_k], Y) = 0$ for each $k \geq 0$. Then

$$c = \lim_{t \rightarrow \infty} \operatorname{Dist}(z[t_k], Y) = 0,$$

contradicting the assumption that $c > 0$.

Since $z[t] \notin Y$ for each $t \geq m$, at least one of the following two statements is true.

- (A.1) There exists a subsequence $\{z[t_k]\}_{k=0}^\infty$ such that $z[t_k] < \min Y$ for all $k \geq 0$.
- (A.2) There exists a subsequence $\{z[t'_k]\}_{k=0}^\infty$ such that $z[t'_k] > \max Y$ for all $k \geq 0$.

By symmetry, WLOG, assume (A.1) is true. Then, for each $y \in Y$ and each $k \geq 0$, we have

$$z[t_k] < \min Y \leq y.$$

Thus,

$$|z[t_k] - y| = y - z[t_k].$$

Minimizing over $y \in Y$, we have

$$\text{Dist}(z[t_k], Y) = \min_{y \in Y} |z[t_k] - y| = \min_{y \in Y} (y - z[t_k]) = \min Y - z[t_k].$$

Thus,

$$z[t_k] = \min Y - \text{Dist}(z[t_k], Y). \quad (20)$$

Recall that the limit of $\text{Dist}(z[t], Y)$ exists and $\lim_{t \rightarrow \infty} \text{Dist}(z[t], Y) = c$, and note that $\{\text{Dist}(z[t_k], Y)\}_{k=0}^\infty$ is a subsequence of $\{\text{Dist}(z[t], Y)\}_{t=0}^\infty$. Thus, the limit of $\text{Dist}(z[t_k], Y)$ exists, and

$$\lim_{k \rightarrow \infty} \text{Dist}(z[t_k], Y) = \lim_{t \rightarrow \infty} \text{Dist}(z[t], Y) = c.$$

Therefore, the limit of $z[t_k]$ exists, and

$$\begin{aligned} \lim_{k \rightarrow \infty} z[t_k] &= \lim_{k \rightarrow \infty} (\min Y - \text{Dist}(z[t_k], Y)) \\ &= \min Y - \lim_{k \rightarrow \infty} \text{Dist}(z[t_k], Y) \\ &= \min Y - c. \end{aligned} \quad (21)$$

Thus, $(\min Y - c)$ is an accumulation point of $\{z[t]\}_{t=0}^\infty$.

Similarly, if (A.2) is true, i.e., there exists a subsequence $\{z[t'_k]\}_{k=0}^\infty$ such that $z[t'_k] > \max Y$ for all $k \geq 0$, and we can show that $(\max Y + c)$ is an accumulation point of $\{z[t]\}_{t=0}^\infty$.

Therefore, Lemma 5 has been proved. □

Recall that $\{z[t]\}_{t=0}^\infty$ is a sequence of estimates such that (19) holds, and that there is a sequence of agents $\{j_t\}_{t=0}^\infty$ associated with the sequence $\{z[t]\}_{t=0}^\infty$.

Lemma 6. *If*

$$\lim_{t \rightarrow \infty} \text{Dist}(z[t], Y) = 0, \quad (22)$$

then for each non-faulty agent i in \mathcal{N} , the sequence $\{\text{Dist}(x_i[t], Y)\}_{t=0}^\infty$ converges and

$$\lim_{t \rightarrow \infty} \text{Dist}(x_i[t], Y) = 0.$$

Proof. For each $i \in \mathcal{N}$, we have

$$\text{Dist}(x_i[t], Y) \leq \max_{j \in \mathcal{N}[t]} \text{Dist}(x_j[t], Y) = \text{Dist}(z[t], Y) \quad \text{by (19)}$$

Taking limit sup on both sides, we get

$$\limsup_{t \rightarrow \infty} \text{Dist}(x_i[t], Y) \leq \limsup_{t \rightarrow \infty} \text{Dist}(z[t], Y) = 0 \quad \text{by (22)}$$

Thus, for each $i \in \mathcal{N}$, the sequence $\{\text{Dist}(x_i[t], Y)\}_{t=0}^{\infty}$ converges and

$$\lim_{t \rightarrow \infty} \text{Dist}(x_i[t], Y) = 0.$$

□

Lemma 5 and Lemma 6 derived in proof of Algorithm 1 apply to Algorithm 2 and Algorithm 3 also.

Theorem 1. *The sequence $\{\text{Dist}(z[t], Y)\}_{t=0}^{\infty}$ converges and*

$$\lim_{t \rightarrow \infty} \text{Dist}(z[t], Y) = 0.$$

Proof. Recall that $\mathcal{N}[t-1]$ is the set of agents that do not crash by the end of iteration $t-1$. There may exists an agent j that crashes during the execution of iteration t . If agent j crashes after performing step 3 in Algorithm 1, then $s_j[t]$ is well-defined. On the contrary, if agent j crashes before step 3 is conducted, then $s_j[t]$ is not well-defined. In this case, we define $\text{Dist}(s_j[t], Y) = 0$ for ease of exposition. With this convention, $\min_{j \in \mathcal{N}[t-1]} \text{Dist}(s_j[t], Y)$ is well-defined.

Let $j'_{t-1} \in \mathcal{N}[t-1]$ such that

$$\max_{j \in \mathcal{N}[t-1]} \text{Dist}(s_j[t], Y) = \text{Dist}(s_{j'_{t-1}}[t], Y). \quad (23)$$

We get

$$\begin{aligned} \text{Dist}(z[t], Y) &= \max_{j \in \mathcal{N}[t]} \text{Dist}(x_j[t], Y) \quad \text{due to (19)} \\ &= \max_{j \in \mathcal{N}[t]} \text{Dist}\left(\frac{1}{|\mathcal{R}_j^2[t-1]|} \sum_{i \in \mathcal{R}_j^2[t-1]} s_i[t], Y\right) \quad \text{by (3)} \\ &\leq \max_{j \in \mathcal{N}[t]} \frac{1}{|\mathcal{R}_j^2[t-1]|} \sum_{i \in \mathcal{R}_j^2[t-1]} \text{Dist}(s_i[t], Y) \quad \text{since } \text{Dist}(\cdot, Y) \text{ is convex} \\ &\leq \max_{j \in \mathcal{N}[t]} \max_{i \in \mathcal{R}_j^2[t-1]} \text{Dist}(s_i[t], Y) \\ &\leq \max_{j \in \mathcal{N}[t-1]} \text{Dist}(s_j[t], Y) \\ &= \text{Dist}(s_{j'_{t-1}}[t], Y) \\ &= \inf_{y \in Y} \left| x_{j'_{t-1}}[t-1] - \frac{\lambda[t-1]}{|\mathcal{R}_{j'_{t-1}}^1[t-1]|} \left(\sum_{i \in \mathcal{R}_{j'_{t-1}}^1[t-1]} h'_i(x_{j'_{t-1}}[t-1]) \right) - y \right| \quad \text{by (2).} \quad (24) \end{aligned}$$

Recall that j'_t is defined as (23). Note that for each $t \geq 0$, there exists a non-faulty agent j'_t such that (24) holds, and there exists a sequence of agents $\{j'_t\}_{t=0}^\infty$. Let $\{x[t]\}_{t=0}^\infty$ be a sequence of estimates such that

$$x[t] = x_{j'_t}[t]. \quad (25)$$

Let $\{g[t]\}_{t=0}^\infty$ be a sequence of gradients such that

$$g[t] = \frac{1}{|\mathcal{R}_{j'_t}^1[t]|} \left(\sum_{i \in \mathcal{R}_{j'_t}^1[t]} h'_i(x_{j'_t}[t]) \right). \quad (26)$$

If $x[t-1] = x_{j'_{t-1}}[t-1]$ is a resilient point with respect to the gradient $g[t-1]$, by Definition 1, we can bound (24) further as follows

$$\text{Dist}(z[t], Y) \leq \inf_{y \in Y} \left| x_{j'_{t-1}}[t-1] - \frac{\lambda[t-1]}{|\mathcal{R}_{j'_{t-1}}^1[t-1]|} \left(\sum_{i \in \mathcal{R}_{j'_{t-1}}^1[t-1]} h'_i(x_{j'_{t-1}}[t-1]) \right) - y \right| \leq \lambda[t-1]L. \quad (27)$$

If $x_{j'_{t-1}}[t-1]$ is not a resilient point with respect to the gradient $g[t-1]$, then from Definition 1, we know that

$$\begin{aligned} B1: & \text{ if } x_{j'_{t-1}}[t-1] \in Y, \text{ then } x_{j'_{t-1}}[t-1] - \frac{\lambda[t-1]}{|\mathcal{R}_{j'_{t-1}}^1[t-1]|} \left(\sum_{i \in \mathcal{R}_{j'_{t-1}}^1[t-1]} h'_i(x_{j'_{t-1}}[t-1]) \right) \in Y, \\ B2: & \text{ if } x_{j'_{t-1}}[t-1] < \min Y, \text{ then } x_{j'_{t-1}}[t-1] - \frac{\lambda[t-1]}{|\mathcal{R}_{j'_{t-1}}^1[t-1]|} \left(\sum_{i \in \mathcal{R}_{j'_{t-1}}^1[t-1]} h'_i(x_{j'_{t-1}}[t-1]) \right) \leq \max Y, \\ B3: & \text{ if } x_{j'_{t-1}}[t-1] > \max Y, \text{ then } x_{j'_{t-1}}[t-1] - \frac{\lambda[t-1]}{|\mathcal{R}_{j'_{t-1}}^1[t-1]|} \left(\sum_{i \in \mathcal{R}_{j'_{t-1}}^1[t-1]} h'_i(x_{j'_{t-1}}[t-1]) \right) \geq \min Y. \end{aligned}$$

We consider two scenarios: scenario 1

$$x_{j'_{t-1}}[t-1] - \frac{\lambda[t-1]}{|\mathcal{R}_{j'_{t-1}}^1[t-1]|} \left(\sum_{i \in \mathcal{R}_{j'_{t-1}}^1[t-1]} h'_i(x_{j'_{t-1}}[t-1]) \right) \in Y,$$

and scenario 2

$$x_{j'_{t-1}}[t-1] - \frac{\lambda[t-1]}{|\mathcal{R}_{j'_{t-1}}^1[t-1]|} \left(\sum_{i \in \mathcal{R}_{j'_{t-1}}^1[t-1]} h'_i(x_{j'_{t-1}}[t-1]) \right) \notin Y.$$

The first scenario can possibly appear in each of $B1$, $B2$, and $B3$. In contrast, the second scenario can only appear in $B2$ and $B3$.

Scenario 1: Assume that

$$x_{j'_{t-1}}[t-1] - \frac{\lambda[t-1]}{|\mathcal{R}_{j'_{t-1}}^1[t-1]|} \left(\sum_{i \in \mathcal{R}_{j'_{t-1}}^1[t-1]} h'_i(x_{j'_{t-1}}[t-1]) \right) \in Y,$$

it holds that

$$\inf_{y \in Y} \left| x_{j'_{t-1}}[t-1] - \frac{\lambda[t-1]}{|\mathcal{R}_{j'_{t-1}}^1[t-1]|} \left(\sum_{i \in \mathcal{R}_{j'_{t-1}}^1[t-1]} h'_i(x_{j'_{t-1}}[t-1]) \right) - y \right| = 0 \leq \text{Dist}(z[t-1], Y).$$

Thus, (24) can be further bounded as

$$\begin{aligned} \text{Dist}(z[t], Y) &\leq \inf_{y \in Y} \left| x_{j'_{t-1}}[t-1] - \frac{\lambda[t-1]}{|\mathcal{R}_{j'_{t-1}}^1[t-1]|} \left(\sum_{i \in \mathcal{R}_{j'_{t-1}}^1[t-1]} h'_i(x_{j'_{t-1}}[t-1]) \right) - y \right| \\ &= 0 \end{aligned} \tag{28}$$

$$\leq \text{Dist}(z[t-1], Y). \tag{29}$$

Scenario 2: Assume that

$$x_{j'_{t-1}}[t-1] - \frac{\lambda[t-1]}{|\mathcal{R}_{j'_{t-1}}^1[t-1]|} \left(\sum_{i \in \mathcal{R}_{j'_{t-1}}^1[t-1]} h'_i(x_{j'_{t-1}}[t-1]) \right) \notin Y = [\min Y, \max Y].$$

As commented earlier, either $B2$ holds or $B3$ holds. In addition, from the assumption of scenario 2, $B2$ and $B3$ can be further refined as follows.

$$B2': x_{j'_{t-1}}[t-1] < \min Y \quad \text{and} \quad x_{j'_{t-1}}[t-1] - \frac{\lambda[t-1]}{|\mathcal{R}_{j'_{t-1}}^1[t-1]|} \left(\sum_{i \in \mathcal{R}_{j'_{t-1}}^1[t-1]} h'_i(x_{j'_{t-1}}[t-1]) \right) < \min Y$$

$$B3': x_{j'_{t-1}}[t-1] > \max Y \quad \text{and} \quad x_{j'_{t-1}}[t-1] - \frac{\lambda[t-1]}{|\mathcal{R}_{j'_{t-1}}^1[t-1]|} \left(\sum_{i \in \mathcal{R}_{j'_{t-1}}^1[t-1]} h'_i(x_{j'_{t-1}}[t-1]) \right) > \max Y$$

Suppose $B2'$ is true. As $x_{j'_{t-1}}[t-1] < \min Y$, and $\frac{1}{|\mathcal{R}_{j'_{t-1}}^1[t-1]|} \left(\sum_{i \in \mathcal{R}_{j'_{t-1}}^1[t-1]} h'_i(x_{j'_{t-1}}[t-1]) \right)$ is the gradient of a valid function at point $x_{j'_{t-1}}[t-1]$, from the definition of set Y , we know that

$$\frac{1}{|\mathcal{R}_{j'_{t-1}}^1[t-1]|} \left(\sum_{i \in \mathcal{R}_{j'_{t-1}}^1[t-1]} h'_i(x_{j'_{t-1}}[t-1]) \right) < 0. \tag{30}$$

In addition, since

$$x_{j'_{t-1}}[t-1] - \frac{\lambda[t-1]}{|\mathcal{R}_{j'_{t-1}}^1[t-1]|} \left(\sum_{i \in \mathcal{R}_{j'_{t-1}}^1[t-1]} h'_i(x_{j'_{t-1}}[t-1]) \right) < \min Y,$$

it holds that for any $y \in Y$

$$\begin{aligned}
& \left| x_{j'_{t-1}}[t-1] - \frac{\lambda[t-1]}{|\mathcal{R}_{j'_{t-1}}^1[t-1]|} \left(\sum_{i \in \mathcal{R}_{j'_{t-1}}^1[t-1]} h'_i(x_{j'_{t-1}}[t-1]) \right) - y \right| \\
&= y - x_{j'_{t-1}}[t-1] + \frac{\lambda[t-1]}{|\mathcal{R}_{j'_{t-1}}^1[t-1]|} \left(\sum_{i \in \mathcal{R}_{j'_{t-1}}^1[t-1]} h'_i(x_{j'_{t-1}}[t-1]) \right) \\
&= |y - x_{j'_{t-1}}[t-1]| + \frac{\lambda[t-1]}{|\mathcal{R}_{j'_{t-1}}^1[t-1]|} \left(\sum_{i \in \mathcal{R}_{j'_{t-1}}^1[t-1]} h'_i(x_{j'_{t-1}}[t-1]) \right) \\
&= |y - x_{j'_{t-1}}[t-1]| - \lambda[t-1] \left| \frac{1}{|\mathcal{R}_{j'_{t-1}}^1[t-1]|} \left(\sum_{i \in \mathcal{R}_{j'_{t-1}}^1[t-1]} h'_i(x_{j'_{t-1}}[t-1]) \right) \right| \quad \text{by (30)} \quad (31)
\end{aligned}$$

Similarly, we can show that (31) still holds for the case when $B3'$ is true. Henceforth, we refer (31) as the relation holds for both $B2'$ and $B3'$, i.e., holds under scenario 2.

Thus, under scenario 2, we can bound (24) as follows

$$\begin{aligned}
Dist(z[t], Y) &\leq \inf_{y \in Y} \left| x_{j'_{t-1}}[t-1] - \frac{\lambda[t-1]}{|\mathcal{R}_{j'_{t-1}}^1[t-1]|} \left(\sum_{i \in \mathcal{R}_{j'_{t-1}}^1[t-1]} h'_i(x_{j'_{t-1}}[t-1]) \right) - y \right| \quad \text{by (24)} \\
&= \inf_{y \in Y} |y - x_{j'_{t-1}}[t-1]| - \lambda[t-1] \left| \frac{1}{|\mathcal{R}_{j'_{t-1}}^1[t-1]|} \left(\sum_{i \in \mathcal{R}_{j'_{t-1}}^1[t-1]} h'_i(x_{j'_{t-1}}[t-1]) \right) \right| \quad \text{by (31)} \\
&= Dist(x_{j'_{t-1}}[t-1], Y) - \lambda[t-1] \left| \frac{1}{|\mathcal{R}_{j'_{t-1}}^1[t-1]|} \left(\sum_{i \in \mathcal{R}_{j'_{t-1}}^1[t-1]} h'_i(x_{j'_{t-1}}[t-1]) \right) \right| \\
&\leq Dist(z[t-1], Y) - \lambda[t-1] \left| \frac{1}{|\mathcal{R}_{j'_{t-1}}^1[t-1]|} \left(\sum_{i \in \mathcal{R}_{j'_{t-1}}^1[t-1]} h'_i(x_{j'_{t-1}}[t-1]) \right) \right| \quad (32) \\
&\leq Dist(z[t-1], Y). \quad (33)
\end{aligned}$$

The last inequality follows from the fact that

$$Dist(x_{j'_{t-1}}[t-1], Y) \leq \max_{j \in \mathcal{N}[t-1]} Dist(x_j[t-1], Y) = Dist(z[t-1], Y).$$

Combining the above analysis for the case when $x_{j'_{t-1}}[t-1]$ is a resilient point or the case when $x_{j'_{t-1}}[t-1]$ is not a resilient point, by (27), (29) and (33), we obtain the following iteration relation

$$\text{Dist}(z[t], Y) \leq \max \{ \lambda[t-1]L, \text{Dist}(z[t-1], Y) \}. \quad (34)$$

Recall from (25) and (26) that $x[t-1] = x_{j'_{t-1}}[t-1]$ and $g[t-1] = \frac{1}{|\mathcal{R}_{j'_{t-1}}^1[t-1]|} \left(\sum_{i \in \mathcal{R}_{j'_{t-1}}^1[t-1]} h'_i(x_{j'_{t-1}}[t-1]) \right)$.

We consider two cases : case (i) there are infinitely many points in $\{x[t]\}_{t=0}^\infty$ that are resilient with respect to $\{g[t]\}_{t=0}^\infty$, and case (ii) there are finitely many points in $\{x[t]\}_{t=0}^\infty$ that are resilient with respect to $\{g[t]\}_{t=0}^\infty$, respectively.

Case (i): There are infinitely many points in $\{x[t]\}_{t=0}^\infty$ that are resilient with respect to $\{g[t]\}_{t=0}^\infty$.

Let $\{t_i\}_{i=0}^\infty$ be the maximal sequence of such indices. Since $x[t_i]$ is a resilient point with respect to $g[t_i]$ for each i , then for each t_i , by (27), we get

$$\text{Dist}(z[t_i+1], Y) \leq \lambda[t_i]L, \quad (35)$$

and for each $t \neq t_i$ for any i , by (29) and (33), we get

$$\text{Dist}(z[t+1], Y) \leq \text{Dist}(z[t], Y). \quad (36)$$

Taking limit sup on both sides of (35) over i , we get

$$0 \leq \limsup_{i \rightarrow \infty} \text{Dist}(z[t_i+1], Y) \leq \limsup_{i \rightarrow \infty} \lambda[t_i]L = 0. \quad (37)$$

For each $\tau > t_0$ and $\tau \notin \{t_i\}_{i=0}^\infty$, there exists $t_{i(\tau)}$ such that $t_{i(\tau)} < \tau \leq t_{i(\tau)+1}$. Then, we get

$$\text{Dist}(z[\tau], Y) \leq \text{Dist}(z[t_{i(\tau)}+1], Y) \quad \text{due to (36) and that } \tau \geq t_{i(\tau)}+1 \quad (38)$$

$$\leq \lambda[t_{i(\tau)}]L \quad \text{by (35)} \quad (39)$$

Taking the limit sup on both sides of (38) over τ , where $\tau > t_0$ and $\tau \notin \{t_i\}_{i=0}^\infty$, we get

$$\limsup_{\tau \rightarrow \infty} \text{Dist}(z[\tau], Y) \leq \limsup_{\tau \rightarrow \infty} \lambda[t_{i(\tau)}]L = \lim_{\tau \rightarrow \infty} \lambda[t_{i(\tau)}]L = 0. \quad (40)$$

From (37), we know that $\forall \epsilon > 0, \exists i_0$ such that for all $i \geq i_0$, the following holds.

$$\sup\{\text{Dist}(z[t_j], Y), t_j \in \{t_i\}_{i=0}^\infty, j \geq i_0\} = |\sup\{\text{Dist}(z[t_j], Y), t_j \in \{t_i\}_{i=0}^\infty, j \geq i_0\} - 0| < \epsilon. \quad (41)$$

From (40), we know that $\forall \epsilon > 0, \exists \tau^*, \tau^* \notin \{t_i\}_{i=0}^\infty$ such that for all $\tau \geq \tau^*, \tau \notin \{t_i\}_{i=0}^\infty$, the following holds.

$$\sup\{\text{Dist}(z[\tau], Y), \tau \geq \tau^*, \tau \notin \{t_i\}_{i=0}^\infty\} = |\sup\{\text{Dist}(z[\tau], Y), \tau \geq \tau^*, \tau \notin \{t_i\}_{i=0}^\infty\} - 0| < \epsilon. \quad (42)$$

Let $t^* = \max\{t_{i_0}, \tau^*\}$. Then for each $\epsilon > 0$ and $t \geq t^*$, we have

$$\begin{aligned} & \sup\{\text{Dist}(z[t], Y), t \geq t^*\} \\ &= \sup\{(\text{Dist}(z[t], Y), t \in \{t_i\}_{i=0}^\infty, t \geq t_{i_0}) \cup (\text{Dist}(z[t], Y), t \notin \{t_i\}_{i=0}^\infty, t \geq \tau^*)\} \\ &= \max\{\sup\{\text{Dist}(z[t], Y), t \in \{t_i\}_{i=0}^\infty, t \geq t_{i_0}\}, \sup\{\text{Dist}(z[t], Y), t \notin \{t_i\}_{i=0}^\infty, t \geq \tau^*\}\} \\ &< \max\{\epsilon, \epsilon\} = \epsilon \quad \text{by (41) and (42)}. \end{aligned}$$

Thus, we have

$$\limsup_{t \rightarrow \infty} \text{Dist}(z[t], Y) = 0.$$

Therefore, the limit of $\text{Dist}(z[t], Y)$ exists, and

$$\lim_{t \rightarrow \infty} \text{Dist}(z[t], Y) = 0.$$

Case (ii): There are finitely many points in $\{x[t]\}_{t=0}^\infty$ that are resilient with respect to $\{g[t]\}_{t=0}^\infty$.

By the assumption of case (ii) we know that there exists a time index m_0 such that for all $t \geq m_0$, each $x[t]$ is not a resilient point with respect to $g[t]$. Then, for $t \geq m_0$, (36) holds. Thus, by MCT, the limit of $\text{Dist}(z[t], Y)$ exists. Let $c \geq 0$ be a nonnegative constant such that

$$\lim_{t \rightarrow \infty} \text{Dist}(z[t], Y) = c. \quad (43)$$

Since $\text{Dist}(z[t], Y) \leq \text{Dist}(z[m_0], Y)$ holds for each $t \geq m_0$, we know that $c < \infty$.

Case (ii.a): Assume that there are infinitely many time indices $t \geq m_0$ such that

$$x[t] - \lambda[t]g[t] = x_{j'_t}[t] - \frac{\lambda[t]}{|\mathcal{R}_{j'_t}^1[t]|} \left(\sum_{i \in \mathcal{R}_{j'_t}^1[t]} h'_i(x_{j'_t}[t]) \right) \in Y.$$

Let $\{t_k\}_{k=0}^\infty$ be the maximal sequence of such time indices. By (28), we have

$$\text{Dist}(z[t_k + 1], Y) \leq 0.$$

Thus, the limit of $\text{Dist}(z[t_k + 1], Y)$ exists and

$$\lim_{k \rightarrow \infty} \text{Dist}(z[t_k + 1], Y) = 0.$$

Recall from (43) that the limit of $\text{Dist}(z[t], Y)$ exists. The limit of $\text{Dist}(z[t], Y)$ and the limit of $\text{Dist}(z[t_k + 1], Y)$ should be identical, i.e.,

$$c = \lim_{t \rightarrow \infty} \text{Dist}(z[t], Y) = \lim_{k \rightarrow \infty} \text{Dist}(z[t_k + 1], Y) = 0, \quad (44)$$

proving the theorem.

Case (ii.b): Assume that there are only finitely many time indices $t \geq m_0$ such that

$$x[t] - \lambda[t]g[t] = x_{j'_t}[t] - \frac{\lambda[t]}{|\mathcal{R}_{j'_t}^1[t]|} \left(\sum_{i \in \mathcal{R}_{j'_t}^1[t]} h'_i(x_{j'_t}[t]) \right) \in Y.$$

Then, there exists² $m' \geq m_0$ such that for each $t \geq m' \geq m_0$, $x[t]$ is not a resilient point with respect to $g[t]$, and $x[t] - \lambda[t]g[t] \notin Y$. Thus, for each $t \geq m' \geq m_0$, (32) holds, i.e.,

$$\text{Dist}(z[t + 1], Y) \leq \text{Dist}(z[t], Y) - \lambda[t] \left| \frac{1}{|\mathcal{R}_{j'_t}^1[t]|} \left(\sum_{i \in \mathcal{R}_{j'_t}^1[t]} h'_i(x_{j'_t}[t]) \right) \right|.$$

² Recall that m_0 is the time index such that for each $t \geq m_0$, $x[t]$ is not a resilient point with respect to $g[t]$.

Recall that $0 \leq c < \infty$ is a nonnegative constant such that $\lim_{t \rightarrow \infty} \text{Dist}(z[t], Y) = c$. **Next we show that $c = 0$.** We prove this by contradiction. Suppose $c > 0$. By Lemma 5, we know that either (A.1) is true or (A.2) is true.

(A.1) There exists a subsequence $\{z[t_k]\}_{k=0}^\infty$ such that $z[t_k] < \min Y$ for all $k \geq 0$.

(A.2) There exists a subsequence $\{z[t'_k]\}_{k=0}^\infty$ such that $z[t'_k] > \max Y$ for all $k \geq 0$.

We also know that at least one of $(\min Y - c)$ or $(\max Y + c)$ is an accumulation point of $\{z[t]\}_{t=0}^\infty$, and no other accumulation points exist.

Let $a = \min Y$, $b = \max Y$ and $\epsilon = \frac{c}{2}$. It can be seen from the proof of Lemma 5 that there exists m such that $z[t] \notin Y$ for each $t \geq m$. We consider three scenarios: (A.1) is true but (A.2) is not true, (A.2) is true but (A.1) is not true, both (A.1) and (A.2) are true.

When (A.1) holds but (A.2) does not hold: That is, there exists a subsequence $\{z[t_k]\}_{k=0}^\infty$ such that $z[t_k] < \min Y$ for all $k \geq 0$; and there does not exist a subsequence $\{z[t'_k]\}_{k=0}^\infty$ such that $z[t'_k] > \max Y$ for all $k \geq 0$. Then there exists $m_1 \geq m$ such that $z[t] < \min Y$ for each $t \geq m_1 \geq m$. From the proof of Lemma 5, we know

$$\lim_{t \rightarrow \infty} z[t] = \min Y - c = a - c. \quad (45)$$

Since (45) holds, there exists $m_1^* \geq m_1 \geq m$ such that for all $t \geq m_1^* \geq m_1 \geq m$, the following holds.

$$|z[t] - (a - c)| \leq \epsilon = \frac{c}{2} \iff a - \frac{3c}{2} \leq z[t] \leq a - \frac{c}{2}. \quad (46)$$

Since $c > 0$, we have $a - \frac{c}{2} < a$. Then, for each $p(\cdot) \in \mathcal{C}$, $p'(a - \frac{c}{2}) < 0$. Thus,

$$\rho^* \triangleq \sup_{p(\cdot) \in \mathcal{C}} p'(a - \frac{c}{2}) \leq 0.$$

Let $K = \sum_{j \in \mathcal{F}} \mathbf{1}\{h'_j(a - \frac{c}{2}) \geq 0\}$. Define $q(x)$ as follows,

$$q(x) = \frac{1}{|\mathcal{N}| + K} \left(\sum_{j \in \mathcal{N}} h_j(x) + \sum_{j \in \mathcal{F}} h_j(x) \mathbf{1}\{h'_j(a - \frac{c}{2}) \geq 0\} \right).$$

It can be easily seen that $q(\cdot) \in \mathcal{C}$ is a valid function and

$$\rho^* = \sup_{p(\cdot) \in \mathcal{C}} p'(a - \frac{c}{2}) = q'(a - \frac{c}{2}) < 0.$$

Note that when $t \geq m_1^* \geq m_1 \geq m$, (32) may not hold, since it is possible that $z[t] - \lambda[t]g[t] \in Y$. Let $\tilde{t}_1 = \max\{m_1^*, m'\}$. For each $t \geq \tilde{t}_1 = \max\{m_1^*, m'\}$, (32), (45) and (46) hold. We have

$$\begin{aligned} \text{Dist}(z[t+1], Y) &\leq \text{Dist}(z[t], Y) - \lambda[t] \left| \frac{1}{|\mathcal{R}_{j'_t}^1[t]|} \left(\sum_{i \in \mathcal{R}_{j'_t}^1[t]} h'_i(x_{j'_t}[t]) \right) \right| \quad \text{by (32)} \\ &= \text{Dist}(z[t], Y) - \lambda[t] \left| \frac{1}{|\mathcal{R}_{j'_t}^1[t]|} \left(\sum_{i \in \mathcal{R}_{j'_t}^1[t]} h'_i(z[t]) \right) \right| + \frac{1}{|\mathcal{R}_{j'_t}^1[t]|} \sum_{i \in \mathcal{R}_{j'_t}^1[t]} \left(h'_i(x_{j'_t}[t]) - h'_i(z[t]) \right) \\ &\stackrel{(a)}{\leq} \text{Dist}(z[t], Y) - \lambda[t] \left| \frac{1}{|\mathcal{R}_{j'_t}^1[t]|} \left(\sum_{i \in \mathcal{R}_{j'_t}^1[t]} h'_i(z[t]) \right) \right| + \lambda[t](M[t] - m[t])L \\ &\leq \text{Dist}(z[t], Y) - \lambda[t]|\rho^*| + \lambda[t](M[t] - m[t])L. \end{aligned} \quad (47)$$

Inequality (a) holds because gradient $h'_k(\cdot)$ is L -Lipschitz for each $k \in \mathcal{V}$,

$$\left| x_{j'_{t+1}}[t] - x_k[t] \right| \leq \max_{i,j \in \mathcal{N}[t]} (x_i[t] - x_j[t]) = \max_{i \in \mathcal{N}[t]} x_i[t] - \min_{j \in \mathcal{N}[t]} x_j[t] = M[t] - m[t],$$

and the fact that

$$\frac{1}{|\mathcal{R}_{j'_t}[t]|} \sum_{k \in \mathcal{R}_{j'_t}[t]} 1 = 1.$$

Next we show that the last inequality holds. Since $h'_i(\cdot)$ is non-decreasing for each $i \in \mathcal{V}$, then the function

$$\frac{1}{|\mathcal{R}_{j'_t}^1[t]|} \left(\sum_{i \in \mathcal{R}_{j'_t}^1[t]} h'_i(\cdot) \right)$$

is non-decreasing. In addition, by (46) we know that $a - \frac{3c}{2} \leq z[t] \leq a - \frac{c}{2}$. We get

$$\frac{1}{|\mathcal{R}_{j'_t}^1[t]|} \left(\sum_{i \in \mathcal{R}_{j'_t}^1[t]} h'_i(z[t]) \right) \leq \frac{1}{|\mathcal{R}_{j'_t}^1[t]|} \left(\sum_{i \in \mathcal{R}_{j'_t}^1[t]} h'_i(a - \frac{c}{2}) \right) \leq \sup_{p(\cdot) \in \mathcal{C}} p'(a - \frac{c}{2}) = \rho^* < 0.$$

Thus,

$$\left| \frac{1}{|\mathcal{R}_{j'_t}^1[t]|} \left(\sum_{i \in \mathcal{R}_{j'_t}^1[t]} h'_i(z[t]) \right) \right| \geq |\rho^*|, \quad (48)$$

proving the last inequality in (47). Repeatedly apply (47) for $t \geq \tilde{t}_1 = \max\{m_1^*, m'\}$, we get

$$\text{Dist}(z[t+1], Y) \leq \text{Dist}(z[\tilde{t}_1], Y) - \left(\sum_{r=\tilde{t}_1}^t \lambda[r] \right) |\rho^*| + \sum_{r=\tilde{t}_1}^t \lambda[r] (M[r] - m[r])L. \quad (49)$$

Taking limit on both sides of (49), we obtain

$$\begin{aligned} \lim_{t \rightarrow \infty} \text{Dist}(z[t+1], Y) &\leq \text{Dist}(z[\tilde{t}_1], Y) - \left(\sum_{r=\tilde{t}_1}^{\infty} \lambda[r] \right) |\rho^*| + \sum_{r=\tilde{t}_1}^{\infty} \lambda[r] (M[r] - m[r])L \\ &\leq \text{Dist}(z[\tilde{t}_1], Y) - \left(\sum_{r=\tilde{t}_1}^{\infty} \lambda[r] \right) |\rho^*| + \sum_{r=0}^{\infty} \lambda[r] (M[r] - m[r])L \\ &\stackrel{(a)}{=} \text{Dist}(z[\tilde{t}_1], Y) - \infty + C_1 \\ &= -\infty. \end{aligned} \quad (50)$$

Equality (a) is true due to (16), the fact that $|\rho^*| > 0$ and that

$$\sum_{r=\tilde{t}_1}^{\infty} \lambda[r] = \sum_{t=0}^{\infty} \lambda[t] - \sum_{r=0}^{\tilde{t}_1-1} \lambda[r] = \infty - \sum_{r=0}^{\tilde{t}_1-1} \lambda[r] = \infty.$$

On the other hand, we know $\lim_{t \rightarrow \infty} \text{Dist}(z[t], Y) = c > 0$. This is a contradiction. Thus,

$$\lim_{t \rightarrow \infty} \text{Dist}(z[t], Y) = c = 0.$$

When (A.2) holds but (A.1) does not hold: That is, there does not exist a subsequence $\{z[t_k]\}_{k=0}^\infty$ such that $z[t_k] < \min Y$ for all $k \geq 0$; and there exists a subsequence $\{z[t'_k]\}_{k=0}^\infty$ such that $z[t'_k] > \max Y$ for all $k \geq 0$. Recall that $z[t] \notin Y$ for each $t \geq m$. Then there exists $m_2 \geq m$ such that $z[t] > \max Y$ for each m_2 . From the proof of Lemma 5, we get

$$\lim_{t \rightarrow \infty} z[t] = \max Y + c = b + c. \quad (51)$$

Since (51) holds, there exists $m_2^* \geq m_2 \geq m$ such that for all $t \geq m_2^* \geq m_2 \geq m$, the following holds.

$$|z[t] - (b + c)| \leq \epsilon = \frac{c}{2} \iff b + \frac{c}{2} \leq z[t] \leq b + \frac{3c}{2}. \quad (52)$$

Since $c > 0$, we have $b + \frac{c}{2} > b$. Then, for each $p(\cdot) \in \mathcal{C}$, $p'(b + \frac{c}{2}) > 0$. Then,

$$\tilde{\rho} \triangleq \inf_{p(\cdot) \in \mathcal{C}} p'(b + \frac{c}{2}) \geq 0.$$

Let $K = \sum_{j \in \mathcal{F}} \mathbf{1}\{h'_j(b + \frac{c}{2}) \leq 0\}$. Define $\tilde{q}(x)$ as follows,

$$\tilde{q}(x) = \frac{1}{|\mathcal{N}| + K} \left(\sum_{j \in \mathcal{N}} h_j(x) + \sum_{j \in \mathcal{F}} h_j(x) \mathbf{1}\{h'_j(b + \frac{c}{2}) \leq 0\} \right).$$

It can be easily seen that $\tilde{q}(\cdot) \in \mathcal{C}$ is a valid function and

$$\inf_{p(\cdot) \in \mathcal{C}} p'(b + \frac{c}{2}) = \tilde{q}'(b + \frac{c}{2}) > 0.$$

Then, $\tilde{\rho} = \tilde{q}'(b + \frac{c}{2}) > 0$.

Note that when $t \geq m_2^* \geq m_2 \geq m$, (32) may not hold, since it is possible that $z[t] - \lambda[t]g[t] \in Y$. Let $\tilde{t}_2 = \max\{m_2^*, m'\}$. For each $t \geq \tilde{t}_2 = \max\{m_2^*, m'\}$, (32), (51) and (52) hold. We have

$$\begin{aligned} \text{Dist}(z[t+1], Y) &\leq \text{Dist}(z[t], Y) - \lambda[t] \left| \frac{1}{|\mathcal{R}_{j'_t}^1[t]|} \left(\sum_{i \in \mathcal{R}_{j'_t}^1[t]} h'_i(x_{j'_t}[t]) \right) \right| \quad \text{by (32)} \\ &\leq \text{Dist}(z[t], Y) - \lambda[t] \left| \frac{1}{|\mathcal{R}_{j'_t}^1[t]|} \left(\sum_{i \in \mathcal{R}_{j'_t}^1[t]} h'_i(z[t]) \right) \right| + \lambda[t](M[t] - m[t])L \\ &\leq \text{Dist}(z[t], Y) - \lambda[t]|\tilde{\rho}| + \lambda[t]L(M[t] - m[t]). \end{aligned} \quad (53)$$

Next we show that the last inequality holds. Recall that the function

$$\frac{1}{|\mathcal{R}_{j'_t}^1[t]|} \left(\sum_{i \in \mathcal{R}_{j'_t}^1[t]} h'_i(\cdot) \right)$$

is non-decreasing. We get

$$\begin{aligned} \frac{1}{|\mathcal{R}_{j'_t}^1[t]|} \left(\sum_{i \in \mathcal{R}_{j'_t}^1[t]} h'_i(z[t]) \right) &\geq \frac{1}{|\mathcal{R}_{j'_t}^1[t]|} \left(\sum_{i \in \mathcal{R}_{j'_t}^1[t]} h'_i(b + \frac{c}{2}) \right) \quad \text{by (52)} \\ &\geq \inf_{p(\cdot) \in \mathcal{C}} p'(b + \frac{c}{2}) = \tilde{\rho} > 0, \end{aligned} \quad (54)$$

proving the last inequality in (53). Repeatedly apply (53) for $t \geq \tilde{t}_2 = \max\{m_2^*, m'\}$, we get

$$\text{Dist}(z[t+1], Y) \leq \text{Dist}(z[\tilde{t}_2], Y) - \left(\sum_{r=\tilde{t}_2}^t \lambda[r] \right) |\tilde{\rho}| + \sum_{r=\tilde{t}_2}^t \lambda[r] L(M[r] - m[r]). \quad (55)$$

Taking limit on both sides of (55), we obtain

$$\begin{aligned} \lim_{t \rightarrow \infty} \text{Dist}(z[t+1], Y) &\leq \text{Dist}(z[\tilde{t}_2], Y) - \left(\sum_{r=\tilde{t}_2}^{\infty} \lambda[r] \right) |\tilde{\rho}| + \sum_{r=\tilde{t}_2}^{\infty} \lambda[r] (M[r] - m[r]) L \\ &\leq \text{Dist}(z[\tilde{t}_2], Y) - \left(\sum_{r=\tilde{t}_2}^{\infty} \lambda[r] \right) |\tilde{\rho}| + \sum_{r=0}^{\infty} \lambda[r] (M[r] - m[r]) L \\ &\leq \text{Dist}(z[\tilde{t}_2], Y) - \infty + C_1 \\ &= -\infty. \end{aligned}$$

This inequality is obtained similarly to the inequality (50). On the other hand, we know $\lim_{t \rightarrow \infty} \text{Dist}(z[t], Y) = c > 0$. This is a contradiction. Thus,

$$\lim_{t \rightarrow \infty} \text{Dist}(z[t], Y) = c = 0.$$

Both (A.1) and (A.2) hold: Let $\{z[t_k]\}_{k=0}^{\infty}$ be a maximal subsequence of $\{z[t]\}_{t=0}^{\infty}$ such that $t_k \geq m'$ and $z[t_k] < \min Y$ for all $k \geq 0$. Let $\{z[t'_k]\}_{k=0}^{\infty}$ be a maximal subsequence of $\{z[t]\}_{t=0}^{\infty}$ such that $t'_k \geq m'$ and $t'_k > \max Y$ for all $k \geq 0$. Recall that $z[t] \notin Y$ for each $t \geq m$. Then,

$$\{z[t_k]\}_{k=0}^{\infty} \cup \{z[t'_k]\}_{k=0}^{\infty} = \{z[t]\}_{t \geq m}.$$

By Lemma 5, we know

$$\lim_{k \rightarrow \infty} z[t_k] = \min Y - c = a - c \quad \text{and} \quad \lim_{k \rightarrow \infty} z[t'_k] = \max Y + c = b + c. \quad (56)$$

Since

$$\{z[t_k]\}_{k=0}^{\infty} \cup \{z[t'_k]\}_{k=0}^{\infty} = \{z[t]\}_{t \geq m},$$

there exist $m_3 \geq m$ such that for each $t \geq m_3$,

$$a - \frac{3c}{2} \leq z[t] \leq a - \frac{c}{2} \quad \text{or} \quad b + \frac{c}{2} \leq z[t] \leq b + \frac{3c}{2}. \quad (57)$$

Recall that

$$\rho^* = \sup_{p(\cdot) \in \mathcal{C}} p'(a - \frac{c}{2}) = q'(a - \frac{c}{2}) \quad \text{and} \quad \tilde{\rho} \triangleq \inf_{p(\cdot) \in \mathcal{C}} p'(b + \frac{c}{2}) = \tilde{q}'(b + \frac{c}{2}).$$

Recall that for each $t \geq m' \geq m_0$, $z[t]$ is not a resilient point with respect to $g[t]$, and $z[t] - \lambda[t]g[t] \notin Y$. Thus (32) holds, i.e.,

$$\text{Dist}(z[t+1], Y) \leq \text{Dist}(z[t], Y) - \lambda[t] \left| \frac{1}{|\mathcal{R}_{j'_t}^1[t]|} \left(\sum_{i \in \mathcal{R}_{j'_t}^1[t]} h'_i(x_{j'_t}[t]) \right) \right|.$$

Since (57), by (48) and (54), we get

$$\left| \frac{1}{|\mathcal{R}_{j'_t}^1[t]|} \left(\sum_{i \in \mathcal{R}_{j'_t}^1[t]} h'_i(z[t]) \right) \right| \geq |\rho^*| \quad \text{or} \quad \left| \frac{1}{|\mathcal{R}_{j'_t}^1[t]|} \left(\sum_{i \in \mathcal{R}_{j'_t}^1[t]} h'_i(z[t]) \right) \right| \geq |\tilde{\rho}|.$$

Thus,

$$\left| \frac{1}{|\mathcal{R}_{j'_t}^1[t]|} \left(\sum_{i \in \mathcal{R}_{j'_t}^1[t]} h'_i(z[t]) \right) \right| \geq \min\{|\rho^*|, |\tilde{\rho}|\}. \quad (58)$$

We get

$$\begin{aligned} \text{Dist}(z[t+1], Y) &\leq \text{Dist}(z[t], Y) - \lambda[t] \left| \frac{1}{|\mathcal{R}_{j'_t}^1[t]|} \left(\sum_{i \in \mathcal{R}_{j'_t}^1[t]} h'_i(x_{j'_t}[t]) \right) \right| \quad \text{by (32)} \\ &\leq \text{Dist}(z[t], Y) - \lambda[t] \left| \frac{1}{|\mathcal{R}_{j'_t}^1[t]|} \left(\sum_{i \in \mathcal{R}_{j'_t}^1[t]} h'_i(z[t]) \right) \right| + \lambda[t](M[t] - m[t])L \\ &\leq \text{Dist}(z[t], Y) - \lambda[t] \min\{|\rho^*|, |\tilde{\rho}|\} + \lambda[t]L(M[t] - m[t]) \quad \text{by (58)} \end{aligned} \quad (59)$$

Repeatedly apply (59) for $t \geq \tilde{t}_3 = \max\{m_3^*, m'\}$, we get

$$\text{Dist}(z[t+1], Y) \leq \text{Dist}(z[\tilde{t}_3], Y) - \left(\sum_{r=\tilde{t}_3}^t \lambda[r] \right) \min\{|\rho^*|, |\tilde{\rho}|\} + \sum_{r=\tilde{t}_3}^t \lambda[r]L(M[r] - m[r]). \quad (60)$$

Taking limit on both sides of (59), we obtain

$$\begin{aligned} \lim_{t \rightarrow \infty} \text{Dist}(z[t+1], Y) &\leq \text{Dist}(z[\tilde{t}_3], Y) - \left(\sum_{r=\tilde{t}_3}^{\infty} \lambda[r] \right) \min\{|\rho^*|, |\tilde{\rho}|\} + \sum_{r=\tilde{t}_3}^{\infty} \lambda[r]L(M[r] - m[r]) \\ &\leq \text{Dist}(z[\tilde{t}_3], Y) - \infty + C_1 \\ &= -\infty. \end{aligned}$$

This inequality is obtained similarly to the inequality (50).

On the other hand, we know $\lim_{t \rightarrow \infty} \text{Dist}(z[t], Y) = c > 0$. A contradiction is proved. Thus,

$$\lim_{t \rightarrow \infty} \text{Dist}(z[t], Y) = c = 0.$$

The proof is complete. □

3.2 Algorithm 2

In Algorithm 1, in each iteration $t \geq 1$, there are two rounds of information exchange. Next we will present a simple algorithm which only requires one message sent by each agent per iteration. In this algorithm, each agent j maintains one local estimate x_j , where $x_j[0]$ is an arbitrary input at agent j .

Algorithm 2 for agent j for iteration $t \geq 1$:

Step 1: Compute $h'_j(x_j[t-1])$ —the gradient of local function $h_j(\cdot)$ at point $x_j[t-1]$, and send the tuple $(x_j[t-1], h'_j(x_j[t-1]))$ to all the agents (including agent j itself).

Step 2: Let $R_j[t-1]$ denote the set of tuples of the form $(x_i[t-1], h'_i(x_i[t-1]))$ received as a result of step 1. Update x_j as

$$x_j[t] = \frac{1}{|R_j[t-1]|} \left(\sum_{i \in R_j[t-1]} (x_i[t-1] - \lambda[t-1] h'_i(x_i[t-1])) \right). \quad (61)$$

Note that $R_j[t-1] \subseteq \mathcal{N}[t-1]$. In addition, set Y is the same as that defined earlier for Algorithm 1.

Lemma 7. *Under Algorithm 2, the sequence $\{M[t] - m[t]\}_{t=0}^\infty$ converges and*

$$\lim_{t \rightarrow \infty} (M[t] - m[t]) = 0.$$

Recall that $M[t] = \max_{i \in \mathcal{N}[t]} x_i[t]$ and $m[t] = \min_{i \in \mathcal{N}[t]} x_i[t]$. Lemma 7 implies that asymptotic consensus is achieved under Algorithm 2. The proof of Lemma 7 is similar to the proof of Lemma 3, and is omitted.

Lemma 8. *Under Algorithm 2, the following holds.*

$$\sum_{t=0}^{\infty} \lambda[t] (M[t] - m[t]) < \infty.$$

The proof of Lemma 8 is the similar to the proof of Lemma 4, and is omitted. By Lemma 8, we know there exists some constant C_2 such that for any constant $t \geq 0$,

$$\sum_{\tau=t}^{\infty} \lambda[\tau] L(M[\tau] - m[\tau]) \leq \sum_{\tau=0}^{\infty} \lambda[\tau] L(M[\tau] - m[\tau]) \leq C_2. \quad (62)$$

The following corollary is an immediate consequence of Lemma 8.

Corollary 2. *Under Algorithm 2,*

$$\lim_{t \rightarrow \infty} \lambda[t] (M[t] - m[t]) = 0,$$

and

$$\lim_{t \rightarrow \infty} \sum_{\tau=t}^{\infty} \lambda[\tau] (M[\tau] - m[\tau]) = 0.$$

The proof of Corollary 2 is similar to the proof of Corollary 1, and is omitted.

In our convergence analysis, we will use the well-know “almost supermartingale” convergence theorem in [18], which can also be found as Lemma 11, in Chapter 2.2 [17]. We present a simpler deterministic version of the theorem in the next lemma.

Lemma 9. [18] *Let $\{a_t\}_{t=0}^\infty$, $\{b_t\}_{t=0}^\infty$, and $\{c_t\}_{t=0}^\infty$ be non-negative sequences. Suppose that*

$$a_{t+1} \leq a_t - b_t + c_t \quad \text{for all } t \geq 0,$$

and $\sum_{t=0}^\infty c_t < \infty$. Then $\sum_{t=0}^\infty b_t < \infty$ and the sequence $\{a_t\}_{t=0}^\infty$ converges to a non-negative value.

Recall that set Y is the same as that defined earlier for Algorithm 1. We define $z[t]$ and x_{j_t} similar to that for Algorithm 1. In particular, let $\{z[t]\}_{t=0}^\infty$ be a sequence of estimates such that

$$z[t] = x_{j_t}[t], \quad \text{where } j_t \in \operatorname{argmax}_{j \in \mathcal{N}[t]} \operatorname{Dist}(x_j[t], Y). \quad (63)$$

From the definition, there is a sequence of agents $\{j_t\}_{t=0}^\infty$ associated with the sequence $\{z[t]\}_{t=0}^\infty$.

Theorem 2. *The sequence $\{\operatorname{Dist}(z[t], Y)\}_{t=0}^\infty$ converges and*

$$\lim_{t \rightarrow \infty} \operatorname{Dist}(z[t], Y) = 0.$$

Proof. We first try to derive an iteration relation similar to that in (34).

$$\begin{aligned} \operatorname{Dist}(z[t+1], Y) &= \operatorname{Dist}(x_{j_{t+1}}[t+1], Y) \quad \text{by (63)} \\ &= \operatorname{Dist}\left(\frac{1}{|\mathcal{R}_{j_{t+1}}[t]|} \sum_{i \in \mathcal{R}_{j_{t+1}}[t]} (x_i[t] - \lambda[t] h'_i(x_i[t])), Y\right) \quad \text{by (61)} \\ &= \operatorname{Dist}\left(\frac{1}{|\mathcal{R}_{j_{t+1}}[t]|} \sum_{i \in \mathcal{R}_{j_{t+1}}[t]} \left(x_i[t] - \lambda[t] \frac{1}{|\mathcal{R}_{j_{t+1}}[t]|} \sum_{k \in \mathcal{R}_{j_{t+1}}[t]} h'_k(x_k[t])\right), Y\right) \\ &\leq \frac{1}{|\mathcal{R}_{j_{t+1}}[t]|} \sum_{i \in \mathcal{R}_{j_{t+1}}[t]} \operatorname{Dist}\left(x_i[t] - \lambda[t] \frac{1}{|\mathcal{R}_{j_{t+1}}[t]|} \sum_{k \in \mathcal{R}_{j_{t+1}}[t]} h'_k(x_k[t]), Y\right) \quad \text{by convexity of } \operatorname{Dist}(\cdot, Y) \\ &\leq \max_{i \in \mathcal{R}_{j_{t+1}}[t]} \operatorname{Dist}\left(x_i[t] - \lambda[t] \frac{1}{|\mathcal{R}_{j_{t+1}}[t]|} \sum_{k \in \mathcal{R}_{j_{t+1}}[t]} h'_k(x_k[t]), Y\right) \end{aligned} \quad (64)$$

Let

$$j'_{t+1} \in \operatorname{argmax}_{i \in \mathcal{R}_{j_{t+1}}[t]} \operatorname{Dist}\left(x_i[t] - \lambda[t] \frac{1}{|\mathcal{R}_{j_{t+1}}[t]|} \sum_{k \in \mathcal{R}_{j_{t+1}}[t]} h'_k(x_k[t]), Y\right). \quad (65)$$

Note that $j'_{t+1} \in \mathcal{R}_{j_{t+1}}[t] \subseteq \mathcal{N}[t]$, i.e., $j'_{t+1} \in \mathcal{N}[t]$.

We get

$$\begin{aligned}
Dist(z[t+1], Y) &\leq \max_{i \in \mathcal{R}_{j_{t+1}}[t]} Dist \left(x_i[t] - \lambda[t] \frac{1}{|\mathcal{R}_{j_{t+1}}[t]|} \sum_{k \in \mathcal{R}_{j_{t+1}}[t]} h'_k(x_k[t]), Y \right) \quad \text{by (64)} \\
&= Dist \left(x_{j'_{t+1}}[t] - \lambda[t] \frac{1}{|\mathcal{R}_{j_{t+1}}[t]|} \sum_{k \in \mathcal{R}_{j_{t+1}}[t]} h'_k(x_k[t]), Y \right) \quad \text{by (65)} \\
&= \inf_{y \in Y} \left| x_{j'_{t+1}}[t] - \lambda[t] \frac{1}{|\mathcal{R}_{j_{t+1}}[t]|} \sum_{k \in \mathcal{R}_{j_{t+1}}[t]} h'_k(x_k[t]) - y \right|. \tag{66}
\end{aligned}$$

For each $y \in Y$, we have

$$\begin{aligned}
&\left| x_{j'_{t+1}}[t] - \lambda[t] \frac{1}{|\mathcal{R}_{j_{t+1}}[t]|} \sum_{k \in \mathcal{R}_{j_{t+1}}[t]} h'_k(x_k[t]) - y \right| \\
&= \left| x_{j'_{t+1}}[t] - \lambda[t] \frac{1}{|\mathcal{R}_{j_{t+1}}[t]|} \sum_{k \in \mathcal{R}_{j_{t+1}}[t]} h'_k(x_{j'_{t+1}}[t]) - y + \lambda[t] \frac{1}{|\mathcal{R}_{j_{t+1}}[t]|} \sum_{k \in \mathcal{R}_{j_{t+1}}[t]} (h'_k(x_{j'_{t+1}}[t]) - h'_k(x_k[t])) \right| \\
&\leq \left| x_{j'_{t+1}}[t] - \lambda[t] \frac{1}{|\mathcal{R}_{j_{t+1}}[t]|} \sum_{k \in \mathcal{R}_{j_{t+1}}[t]} h'_k(x_{j'_{t+1}}[t]) - y \right| + \left| \lambda[t] \frac{1}{|\mathcal{R}_{j_{t+1}}[t]|} \sum_{k \in \mathcal{R}_{j_{t+1}}[t]} (h'_k(x_{j'_{t+1}}[t]) - h'_k(x_k[t])) \right| \\
&\leq \left| x_{j'_{t+1}}[t] - \lambda[t] \frac{1}{|\mathcal{R}_{j_{t+1}}[t]|} \sum_{k \in \mathcal{R}_{j_{t+1}}[t]} h'_k(x_{j'_{t+1}}[t]) - y \right| + \lambda[t] \frac{1}{|\mathcal{R}_{j_{t+1}}[t]|} \sum_{k \in \mathcal{R}_{j_{t+1}}[t]} |h'_k(x_{j'_{t+1}}[t]) - h'_k(x_k[t])| \\
&\stackrel{(a)}{\leq} \left| x_{j'_{t+1}}[t] - \lambda[t] \frac{1}{|\mathcal{R}_{j_{t+1}}[t]|} \sum_{k \in \mathcal{R}_{j_{t+1}}[t]} h'_k(x_{j'_{t+1}}[t]) - y \right| + \lambda[t] \frac{1}{|\mathcal{R}_{j_{t+1}}[t]|} \sum_{k \in \mathcal{R}_{j_{t+1}}[t]} L |x_{j'_{t+1}}[t] - x_k[t]| \\
&\stackrel{(b)}{\leq} \left| x_{j'_{t+1}}[t] - \lambda[t] \frac{1}{|\mathcal{R}_{j_{t+1}}[t]|} \sum_{k \in \mathcal{R}_{j_{t+1}}[t]} h'_k(x_{j'_{t+1}}[t]) - y \right| + \lambda[t] L (M[t] - m[t]). \tag{67}
\end{aligned}$$

Inequality (a) holds because gradient $h'_k(\cdot)$ is L -Lipschitz for each $k \in \mathcal{V}$. Inequality (b) holds from the fact that

$$\left| x_{j'_{t+1}}[t] - x_k[t] \right| \leq \max_{i, j \in \mathcal{N}[t]} (x_i[t] - x_j[t]) = \max_{i \in \mathcal{N}[t]} x_i[t] - \min_{j \in \mathcal{N}[t]} x_j[t] = M[t] - m[t],$$

and that

$$\frac{1}{|\mathcal{R}_{j_{t+1}}[t]|} \sum_{k \in \mathcal{R}_{j_{t+1}}[t]} 1 = 1.$$

Using (67), the inequality (66) can be further bounded as

$$\begin{aligned}
Dist(z[t+1], Y) &\leq \inf_{y \in Y} \left| x_{j'_{t+1}}[t] - \lambda[t] \frac{1}{|\mathcal{R}_{j_{t+1}}[t]|} \sum_{k \in \mathcal{R}_{j_{t+1}}[t]} h'_k(x_k[t]) - y \right| \\
&\leq \inf_{y \in Y} \left| x_{j'_{t+1}}[t] - \lambda[t] \frac{1}{|\mathcal{R}_{j_{t+1}}[t]|} \sum_{k \in \mathcal{R}_{j_{t+1}}[t]} h'_k(x_{j'_{t+1}}[t]) - y \right| + \lambda[t] L(M[t] - m[t]) \quad \text{by (67)}.
\end{aligned} \tag{68}$$

Note that for each $t \geq 0$, there exists a non-faulty agent j'_{t+1} such that (67) holds, and there exists a sequence of agents $\{j'_t\}_{t=1}^\infty$. Let $\{x[t]\}_{t=0}^\infty$ be a sequence of estimates such that

$$x[t] = x_{j'_{t+1}}[t]. \tag{69}$$

Let $\{g[t]\}_{t=0}^\infty$ be a sequence of gradients such that

$$g[t] = \frac{1}{|\mathcal{R}_{j_{t+1}}[t]|} \sum_{k \in \mathcal{R}_{j_{t+1}}[t]} h'_k(x_{j'_{t+1}}[t]). \tag{70}$$

If $x[t] = x_{j'_{t+1}}[t]$ is a resilient point with respect to the gradient $g[t] = \frac{1}{|\mathcal{R}_{j_{t+1}}[t]|} \sum_{k \in \mathcal{R}_{j_{t+1}}[t]} h'_k(x_{j'_{t+1}}[t])$, by Definition 1, we bound (68) further as

$$\begin{aligned}
Dist(z[t+1], Y) &\leq \inf_{y \in Y} \left| x_{j'_{t+1}}[t] - \lambda[t] \frac{1}{|\mathcal{R}_{j_{t+1}}[t]|} \sum_{k \in \mathcal{R}_{j_{t+1}}[t]} h'_k(x_{j'_{t+1}}[t]) - y \right| + L\lambda[t] (M[t] - m[t]) \\
&\leq L\lambda[t] + L\lambda[t] (M[t] - m[t]).
\end{aligned} \tag{71}$$

If $x[t] = x_{j'_{t+1}}[t]$ is not a resilient point with respect to the gradient $g[t] = \frac{1}{|\mathcal{R}_{j_{t+1}}[t]|} \sum_{k \in \mathcal{R}_{j_{t+1}}[t]} h'_k(x_{j'_{t+1}}[t])$, then from Definition 1, we know that

- C1: if $x_{j'_{t+1}}[t] \in Y$, then $x_{j'_{t+1}}[t] - \frac{\lambda[t]}{|\mathcal{R}_{j_{t+1}}[t]|} \sum_{k \in \mathcal{R}_{j_{t+1}}[t]} h'_k(x_{j'_{t+1}}[t]) \in Y$,
- C2: if $x_{j'_{t+1}}[t] < \min Y$, then $x_{j'_{t+1}}[t] - \frac{\lambda[t]}{|\mathcal{R}_{j_{t+1}}[t]|} \sum_{k \in \mathcal{R}_{j_{t+1}}[t]} h'_k(x_{j'_{t+1}}[t]) \leq \max Y$,
- C3: if $x_{j'_{t+1}}[t] > \max Y$, then $x_{j'_{t+1}}[t] - \frac{\lambda[t]}{|\mathcal{R}_{j_{t+1}}[t]|} \sum_{k \in \mathcal{R}_{j_{t+1}}[t]} h'_k(x_{j'_{t+1}}[t]) \geq \min Y$.

We consider two scenarios: scenario 1

$$x_{j'_{t+1}}[t] - \frac{\lambda[t]}{|\mathcal{R}_{j_{t+1}}[t]|} \sum_{k \in \mathcal{R}_{j_{t+1}}[t]} h'_k(x_{j'_{t+1}}[t]) \in Y,$$

and scenario 2

$$x_{j'_{t+1}}[t] - \frac{\lambda[t]}{|\mathcal{R}_{j_{t+1}}[t]|} \sum_{k \in \mathcal{R}_{j_{t+1}}[t]} h'_k(x_{j'_{t+1}}[t]) \notin Y.$$

The first scenario can possibly appear in each of C1, C2, and C3. In contrast, the second scenario can only appear in C2 and C3.

Scenario 1: Assume that

$$x_{j'_{t+1}}[t] - \frac{\lambda[t]}{|\mathcal{R}_{j_{t+1}}[t]|} \sum_{k \in \mathcal{R}_{j_{t+1}}[t]} h'_k(x_{j'_{t+1}}[t]) \in Y,$$

it holds that

$$\inf_{y \in Y} \left| x_{j'_{t+1}}[t] - \frac{\lambda[t]}{|\mathcal{R}_{j_{t+1}}[t]|} \sum_{k \in \mathcal{R}_{j_{t+1}}[t]} h'_k(x_{j'_{t+1}}[t]) - y \right| = 0 \leq \text{Dist}(z[t], Y).$$

Thus, (68) can be bounded as

$$\begin{aligned} \text{Dist}(z[t+1], Y) &\leq \inf_{y \in Y} \left| x_{j'_{t+1}}[t] - \lambda[t] \frac{1}{|\mathcal{R}_{j_{t+1}}[t]|} \sum_{k \in \mathcal{R}_{j_{t+1}}[t]} h'_k(x_{j'_{t+1}}[t]) - y \right| + L\lambda[t] (M[t] - m[t]) \\ &\leq 0 + L\lambda[t] (M[t] - m[t]) \end{aligned} \quad (72)$$

$$\leq \text{Dist}(z[t], Y) + L\lambda[t] (M[t] - m[t]). \quad (73)$$

Scenario 2: Assume that

$$x_{j'_{t+1}}[t] - \frac{\lambda[t]}{|\mathcal{R}_{j_{t+1}}[t]|} \sum_{k \in \mathcal{R}_{j_{t+1}}[t]} h'_k(x_{j'_{t+1}}[t]) \notin Y = [\min Y, \max Y].$$

As commented earlier, either $C2$ holds or $C3$ holds. In addition, from the assumption of scenario 2, $C2$ and $C3$ can be further refined as follows.

$$C2': x_{j'_{t+1}}[t] < \min Y \quad \text{and} \quad x_{j'_{t+1}}[t] - \frac{\lambda[t]}{|\mathcal{R}_{j_{t+1}}[t]|} \sum_{k \in \mathcal{R}_{j_{t+1}}[t]} h'_k(x_{j'_{t+1}}[t]) < \min Y$$

$$C3': x_{j'_{t+1}}[t] > \max Y \quad \text{and} \quad x_{j'_{t+1}}[t] - \frac{\lambda[t]}{|\mathcal{R}_{j_{t+1}}[t]|} \sum_{k \in \mathcal{R}_{j_{t+1}}[t]} h'_k(x_{j'_{t+1}}[t]) > \max Y$$

Similar to (30), it can be shown that for both $C2'$ and $C3'$, the following holds.

$$\left| x_{j'_{t+1}}[t] - \frac{\lambda[t]}{|\mathcal{R}_{j_{t+1}}[t]|} \sum_{k \in \mathcal{R}_{j_{t+1}}[t]} h'_k(x_{j'_{t+1}}[t]) - y \right| = |x_{j'_{t+1}}[t] - y| - \lambda[t] \left| \frac{1}{|\mathcal{R}_{j_{t+1}}[t]|} \sum_{k \in \mathcal{R}_{j_{t+1}}[t]} h'_k(x_{j'_{t+1}}[t]) \right|. \quad (74)$$

Thus, under scenario 2, we can bound (68) as

$$\text{Dist}(z[t+1], Y) \leq \inf_{y \in Y} \left| x_{j'_{t+1}}[t] - \lambda[t] \frac{1}{|\mathcal{R}_{j_{t+1}}[t]|} \sum_{k \in \mathcal{R}_{j_{t+1}}[t]} h'_k(x_{j'_{t+1}}[t]) - y \right| + L\lambda[t] (M[t] - m[t])$$

$$= \inf_{y \in Y} |x_{j'_{t+1}}[t] - y| - \lambda[t] \left| \frac{1}{|\mathcal{R}_{j_{t+1}}[t]|} \sum_{k \in \mathcal{R}_{j_{t+1}}[t]} h'_k(x_{j'_{t+1}}[t]) \right| + L\lambda[t] (M[t] - m[t]) \quad \text{by (74)}$$

$$= \text{Dist}(x_{j'_{t+1}}[t], Y) - \lambda[t] \left| \frac{1}{|\mathcal{R}_{j_{t+1}}[t]|} \sum_{k \in \mathcal{R}_{j_{t+1}}[t]} h'_k(x_{j'_{t+1}}[t]) \right| + L\lambda[t] (M[t] - m[t])$$

$$\leq \text{Dist}(z[t], Y) - \lambda[t] \left| \frac{1}{|\mathcal{R}_{j_{t+1}}[t]|} \sum_{k \in \mathcal{R}_{j_{t+1}}[t]} h'_k(x_{j'_{t+1}}[t]) \right| + L\lambda[t] (M[t] - m[t]) \quad (75)$$

$$\leq \text{Dist}(z[t], Y) + L\lambda[t] (M[t] - m[t]). \quad (76)$$

By (71), (73) and (76), for each $t \geq 0$, we obtain the following iteration relation

$$\text{Dist}(z[t+1], Y) \leq \max\{\lambda[t]L, \text{Dist}(z[t], Y)\} + \lambda[t]L(M[t] - m[t]). \quad (77)$$

Recall (69), (70) that $x[t] = x_{j'_{t+1}}[t]$ and $g[t] = \frac{1}{|\mathcal{R}_{j'_{t+1}}[t]|} \sum_{k \in \mathcal{R}_{j'_{t+1}}[t]} h'_k(x_{j'_{t+1}}[t])$. Similar to the proof of Theorem 1, we consider two cases : case (i) there are infinitely many points in $\{x[t]\}_{t=0}^\infty$ that are resilient with respect to $\{g[t]\}_{t=0}^\infty$, and case (ii) there are finitely many points in $\{x[t]\}_{t=0}^\infty$ that are resilient with respect to $\{g[t]\}_{t=0}^\infty$, respectively.

Case (i): There are infinitely many points in $\{x[t]\}_{t=0}^\infty$ that are resilient with respect to $\{g[t]\}_{t=0}^\infty$.

Let $\{t_i\}_{i=0}^\infty$ be the maximal sequence of such indices. Since $x[t_i]$ is a resilient point with respect to $g[t]$ for each i , then for each t_i , by (71), we have

$$\text{Dist}(z[t_i+1], Y) \leq \lambda[t_i]L + \lambda[t_i]L(M[t_i] - m[t_i]), \quad (78)$$

and for each $t \neq t_i \forall i$, by (73) and (76), we get

$$\text{Dist}(z[t+1], Y) \leq \text{Dist}(z[t], Y) + \lambda[t]L(M[t] - m[t]), \quad (79)$$

Taking limit sup on both sides of (78), we get

$$\begin{aligned} \limsup_{i \rightarrow \infty} \text{Dist}(z[t_i+1], Y) &\leq \limsup_{i \rightarrow \infty} \lambda[t_i]L + \limsup_{i \rightarrow \infty} \lambda[t_i]L(M[t_i] - m[t_i]) \\ &= 0 + 0 = 0 \quad \text{by Corollary 2.} \end{aligned} \quad (80)$$

In addition, $\liminf_{i \rightarrow \infty} \text{Dist}(z[t_i+1], Y) \geq 0$. Thus, the limit of $\text{Dist}(z[t_i+1], Y)$ exists, and

$$\lim_{i \rightarrow \infty} \text{Dist}(z[t_i+1], Y) = 0.$$

For each $\tau > t_0$ and $\tau \notin \{t_i\}_{i=0}^\infty$, there exists $t_{i(\tau)}$ such that $t_{i(\tau)} < \tau < t_{i(\tau)+1}$. Repeatedly applying (79), we get

$$\begin{aligned} \text{Dist}(z[\tau+1], Y) &\leq \text{Dist}(z[t_{i(\tau)}+1], Y) + \sum_{r=t_{i(\tau)}+1}^{\tau} \lambda[r]L(M[r] - m[r]) \\ &\leq \lambda[t_{i(\tau)}]L + \lambda[t_{i(\tau)}](M[t_{i(\tau)}] - m[t_{i(\tau)}])L + \sum_{r=t_{i(\tau)}+1}^{\tau} \lambda[r](M[r] - m[r])L \quad \text{by (78)} \\ &= \lambda[t_{i(\tau)}]L + \sum_{r=t_{i(\tau)}}^{\tau} \lambda[r](M[r] - m[r])L \\ &\leq \lambda[t_{i(\tau)}]L + \sum_{r=t_{i(\tau)}}^{\infty} \lambda[r](M[r] - m[r])L \quad \text{since } \lambda[r](M[r] - m[r])L \geq 0, \forall r \end{aligned} \quad (81)$$

Taking limit sup on both sides of (81), we get

$$\begin{aligned} \limsup_{\tau \rightarrow \infty} \text{Dist}(z[\tau+1], Y) &\leq \lim_{\tau \rightarrow \infty} \lambda[t_{i(\tau)}]L + \lim_{\tau \rightarrow \infty} \sum_{r=t_{i(\tau)}}^{\infty} \lambda[r](M[r] - m[r])L \\ &= 0 + 0 = 0 \quad \text{by Corollary 2} \end{aligned}$$

To apply Corollary 2 here we have to have $t_{i(\tau)} \rightarrow \infty$ when $\tau \rightarrow \infty$. This is true since there are infinite resilient points.

Using a similar argument used earlier in the proof of Theorem 1, we conclude that $\lim_{t \rightarrow \infty} \text{Dist}(z[t], Y)$ exists and $\lim_{t \rightarrow \infty} \text{Dist}(z[t], Y) = 0$.

Case (ii): There are finitely many points in $\{x[t]\}_{t=0}^\infty$ that are resilient with respect to $\{g[t]\}_{t=0}^\infty$.

By the assumption in case (ii) we know that there exists a time index m_0 such that for all $t \geq m_0$, each $x[t]$ is not a resilient point with respect to $g[t]$. Thus, for $t \geq m_0$, either (73) or (76) holds. Thus, for $t \geq m_0$, we have

$$\text{Dist}(z[t+1], Y) \leq \text{Dist}(z[t], Y) + \lambda[t]L(M[t] - m[t]). \quad (82)$$

Define $\{a_r\}_{r=0}^\infty$, $\{b_r\}_{r=0}^\infty$, and $\{c_r\}_{r=0}^\infty$ as follows.

$$\begin{aligned} a_r &= \text{Dist}(z[m_0 + r], Y), \\ b_r &= 0, \\ c_r &= \lambda[m_0 + r]L(M[m_0 + r] - m[m_0 + r]). \end{aligned}$$

By Lemma 8 and Lemma 9, we know the limit of $\text{Dist}(z[t], Y)$ exists. Let $c \geq 0$ be a nonnegative constant such that

$$\lim_{t \rightarrow \infty} \text{Dist}(z[t], Y) = c. \quad (83)$$

Repeatedly applying (82), we get

$$\begin{aligned} \text{Dist}(z[t+1], Y) &\leq \text{Dist}(z[t], Y) + \lambda[t]L(M[t] - m[t]) \\ &\leq \text{Dist}(z[m_0], Y) + \sum_{r=m_0}^t \lambda[r]L(M[r] - m[r]) \\ &\leq \text{Dist}(z[m_0], Y) + \sum_{r=m_0}^\infty \lambda[r]L(M[r] - m[r]) \\ &\leq \text{Dist}(z[m_0], Y) + \sum_{r=0}^\infty \lambda[r]L(M[r] - m[r]) \\ &\leq \text{Dist}(z[m_0], Y) + C_2 \quad \text{by (62)}. \end{aligned} \quad (84)$$

Thus, by (84), we know that for each $t \geq m_0$

$$\text{Dist}(z[t+1], Y) \leq \text{Dist}(z[m_0], Y) + C_2.$$

Thus,

$$\lim_{t \rightarrow \infty} \text{Dist}(z[t], Y) = c < \infty.$$

Case (ii.a): Assume that there are infinitely many time indices $t \geq m_0$ such that

$$x[t] - \lambda[t]g[t] = x_{j'_{t+1}}[t] - \lambda[t] \frac{1}{|\mathcal{R}_{j'_{t+1}}[t]|} \sum_{k \in \mathcal{R}_{j'_{t+1}}[t]} h'_k(x_{j'_{t+1}}[t]) \in Y.$$

Let $\{t_k\}_{k=0}^\infty$ be the maximal sequence of such indices. By (72), we have

$$\text{Dist}(z[t_k+1], Y) \leq 0 + L\lambda[t_k](M[t_k] - m[t_k]). \quad (85)$$

Taking limit on both sides of (85), we get

$$\begin{aligned} \lim_{k \rightarrow \infty} \text{Dist}(z[t_k + 1], Y) &\leq 0 + L \lim_{k \rightarrow \infty} \lambda[t_k] (M[t_k] - m[t_k]) \\ &= 0 + 0 = 0 \quad \text{by Corollary 2} \end{aligned}$$

On the other hand, $\lim_{k \rightarrow \infty} \text{Dist}(z[t_k + 1], Y) = c \geq 0$. Thus,

$$c = \lim_{t \rightarrow \infty} \text{Dist}(z[t], Y) = \lim_{k \rightarrow \infty} \text{Dist}(z[t_k + 1], Y) = 0,$$

proving the theorem.

Case (ii.b): Assume that there are only finitely many time indices $t \geq m_0$ such that

$$x[t] - \lambda[t]g[t] = x_{j'_{t+1}}[t] - \lambda[t] \frac{1}{|\mathcal{R}_{j'_{t+1}}[t]|} \sum_{k \in \mathcal{R}_{j'_{t+1}}[t]} h'_k(x_{j'_{t+1}}[t]) \in Y.$$

Then, there exists $m' \geq m_0$ such that for each $t \geq m' \geq m_0$, $x[t]$ is not a resilient point with respect to $g[t]$, and

$$x[t] - \lambda[t]g[t] = x_{j'_{t+1}}[t] - \lambda[t] \frac{1}{|\mathcal{R}_{j'_{t+1}}[t]|} \sum_{k \in \mathcal{R}_{j'_{t+1}}[t]} h'_k(x_{j'_{t+1}}[t]) \notin Y.$$

Thus, for each $t \geq m' \geq m_0$, (75) holds, i.e.,

$$\text{Dist}(z[t + 1], Y) \leq \text{Dist}(z[t], Y) - \lambda[t] \left| \frac{1}{|\mathcal{R}_{j'_{t+1}}[t]|} \sum_{k \in \mathcal{R}_{j'_{t+1}}[t]} h'_k(x_{j'_{t+1}}[t]) \right| + L\lambda[t] (M[t] - m[t]).$$

Recall that $0 \leq c < \infty$ is a nonnegative constant such that $\lim_{t \rightarrow \infty} \text{Dist}(z[t], Y) = c$. **Next we show that $c = 0$.** We prove this by contradiction. Suppose $c > 0$. By Lemma 5, we know that either (A.1) is true or (A.2) is true.

(A.1) There exists a subsequence $\{z[t_k]\}_{k=0}^{\infty}$ such that $z[t_k] < \min Y$ for all $k \geq 0$.

(A.2) There exists a subsequence $\{z[t'_k]\}_{k=0}^{\infty}$ such that $z[t'_k] > \max Y$ for all $k \geq 0$.

In addition, at least one of $(\min Y - c)$ or $(\max Y + c)$ is an accumulation point of $\{z[t]\}_{t=0}^{\infty}$, and no other accumulation points exist.

Let $a = \min Y$, $b = \max Y$ and $\epsilon = \frac{c}{2}$. It can be seen from the proof of Lemma 5 that there exists m such that $z[t] \notin Y$ for each $t \geq m$. We consider three scenarios: (A.1) is true but (A.2) is not true, (A.2) is true but (A.1) is not true, both (A.1) and (A.2) are true.

When (A.1) holds but (A.2) does not hold: That is, there exists a subsequence $\{z[t_k]\}_{k=0}^{\infty}$ such that $z[t_k] < \min Y$ for all $k \geq 0$; and there does not exist a subsequence $\{z[t'_k]\}_{k=0}^{\infty}$ such that $z[t'_k] > \max Y$ for all $k \geq 0$. Then there exists $m_1 \geq m$ such that $z[t] < \min Y$ for each $t \geq m_1 \geq m$. From the proof of Lemma 5, we know

$$\lim_{t \rightarrow \infty} z[t] = \min Y - c = a - c.$$

Since (45) holds, there exists $m_1^* \geq m_1 \geq m$ such that for all $t \geq m_1^* \geq m_1 \geq m$, the following holds.

$$|z[t] - (a - c)| \leq \epsilon = \frac{c}{2} \quad \Longleftrightarrow \quad a - \frac{3c}{2} \leq z[t] \leq a - \frac{c}{2}. \quad (86)$$

Since $c > 0$, we have $a - \frac{c}{2} < a$. Then, for each $p(\cdot) \in \mathcal{C}$, $p'(a - \frac{c}{2}) < 0$. Then,

$$\rho^* \triangleq \sup_{p(\cdot) \in \mathcal{C}} p'(a - \frac{c}{2}) \leq 0.$$

Let $K = \sum_{j \in \mathcal{F}} \mathbf{1}\{h'_j(a - \frac{c}{2}) \geq 0\}$. Define $q(x)$ as follows,

$$q(x) = \frac{1}{|\mathcal{N}| + K} \left(\sum_{j \in \mathcal{N}} h_j(x) + \sum_{j \in \mathcal{F}} h_j(x) \mathbf{1}\{h'_j(a - \frac{c}{2}) \geq 0\} \right).$$

It can be easily seen that $q(\cdot) \in \mathcal{C}$ is a valid function and

$$\rho^* = \sup_{p(\cdot) \in \mathcal{C}} p'(a - \frac{c}{2}) = q'(a - \frac{c}{2}) < 0.$$

Note that when $t \geq m_1^* \geq m_1 \geq m$, (75) may not hold, since it is possible that $z[t] - \lambda[t]g[t] \in Y$. Let $\tilde{t}_1 = \max\{m_1^*, m'\}$. For each $t \geq \tilde{t}_1 = \max\{m_1^*, m'\}$, (75), (45) and (46) hold. We have

$$\begin{aligned} \text{Dist}(z[t+1], Y) &\leq \text{Dist}(z[t], Y) - \lambda[t] \left| \frac{1}{|\mathcal{R}_{j_{t+1}}[t]|} \left(\sum_{i \in \mathcal{R}_{j_{t+1}}[t]} h'_i(x_{j'_{t+1}}[t]) \right) \right| + \lambda[t]L(M[t] - m[t]) \quad \text{by (75)} \\ &\leq \text{Dist}(z[t], Y) - \lambda[t] \left| \frac{1}{|\mathcal{R}_{j_{t+1}}[t]|} \left(\sum_{i \in \mathcal{R}_{j_{t+1}}[t]} h'_i(z[t]) \right) \right| + 2\lambda[t]L(M[t] - m[t]) \\ &\leq \text{Dist}(z[t], Y) - \lambda[t]|\rho^*| + 2\lambda[t]L(M[t] - m[t]). \end{aligned} \quad (87)$$

Repeatedly applying (87) for $t \geq \tilde{t}_1 = \max\{m_1^*, m'\}$, we get

$$\text{Dist}(z[t+1], Y) \leq \text{Dist}(z[\tilde{t}_1], Y) - \left(\sum_{r=\tilde{t}_1}^t \lambda[r] \right) |\rho^*| + 2 \sum_{r=\tilde{t}_1}^t \lambda[r]L(M[r] - m[r]). \quad (88)$$

Taking limit on both sides of (88), we obtain

$$\begin{aligned} \lim_{t \rightarrow \infty} \text{Dist}(z[t+1], Y) &\leq \text{Dist}(z[\tilde{t}_1], Y) - \left(\sum_{r=\tilde{t}_1}^{\infty} \lambda[r] \right) |\rho^*| + 2 \sum_{r=\tilde{t}_1}^{\infty} \lambda[r]L(M[r] - m[r]) \\ &\leq \text{Dist}(z[\tilde{t}_1], Y) - \left(\sum_{r=\tilde{t}_1}^{\infty} \lambda[r] \right) |\rho^*| + 2C_2 \quad \text{by (62)} \\ &= \text{Dist}(z[\tilde{t}_1], Y) - \infty + 2C_2 \\ &= -\infty. \end{aligned} \quad (89)$$

On the other hand, we know $\lim_{t \rightarrow \infty} \text{Dist}(z[t], Y) = c > 0$. This is a contradiction. Thus,

$$\lim_{t \rightarrow \infty} \text{Dist}(z[t], Y) = c = 0.$$

Similarly, we can show the case when (A.2) holds but (A.1) does not hold, and the case when both (A.1) and (A.2) hold.

The proof of the theorem is complete. □

4 Synchronous Byzantine Iterative Algorithm

In this section, we present an iterative algorithm, in which each non-faulty agent sends only one message per iteration, and keeps minimal memory across iterations. We assume each local cost function $h_j(\cdot)$ has L -Lipschitz continuous derivative.

Algorithm 3 for agent j for iteration $t \geq 1$:

Step 1: Compute $h'_j(x_j[t-1])$ – the gradient of the local cost function $h_j(\cdot)$ at point $x_j[t-1]$, and send the estimate and gradient pair $(x_j[t-1], h'_j(x_j[t-1]))$ to all the agents (including itself).

Step 2: Let $\mathcal{R}_j[t-1]$ denote the set of tuples of the form $(x_i[t-1], h'_i(x_i[t-1]))$ received as a result of step 1.

In step 2, agent j should be able to receive a tuple $(w_i[t-1], g_i[t-1])$ from each agent $i \in \mathcal{V}$. For non-faulty agent $i \in \mathcal{N}$, $w_i[t-1] = x_i[t-1]$ and $g_i[t-1] = h'_i(x_i[t-1])$. If a faulty agent $k \in \mathcal{F}$ does not send a tuple to agent j , then agent j assumes $(w_k[t-1], g_k[t-1])$ to be some default tuple.³

Step 3: Sort the first entries of the received tuples in $\mathcal{R}_j[t-1]$ in a non-increasing order (breaking ties arbitrarily), and erase the smallest f values and the largest f values. Let $\mathcal{R}_j^1[t-1]$ be the identifiers of the $n - 2f$ agents from whom the remaining first entries were received. Similarly, sort the second entries of the received tuples in $\mathcal{R}_j[t-1]$ in a non-increasing order (breaking ties arbitrarily), and erase the smallest f values and the largest f values. Let $\mathcal{R}_j^2[t-1]$ be the identifiers of the $n - 2f$ agents from whom the remaining second entries were received. Denote the largest and smallest gradients among the remaining values by $\hat{g}_j[t-1]$ and $\check{g}_j[t-1]$, respectively. Set $\tilde{g}_j[t-1] = \frac{1}{2}(\hat{g}_j[t-1] + \check{g}_j[t-1])$. Update its state as follows.

$$x_j[t] = \frac{1}{n - 2f} \left(\sum_{i \in \mathcal{R}_j^1[t-1]} w_i[t-1] \right) - \lambda[t-1] \tilde{g}_j[t-1]. \quad (90)$$

Let $\tilde{\mathcal{C}}$ be the collection of functions defined as follows:

$$\begin{aligned} \tilde{\mathcal{C}} \triangleq \{ \quad & p(x) : p(x) = \sum_{i \in \mathcal{N}} \alpha_i h_i(x), \quad \forall i \in \mathcal{N}, \alpha_i \geq 0, \\ & \sum_{i \in \mathcal{N}} \alpha_i = 1, \quad \text{and} \\ & \sum_{i \in \mathcal{N}} \mathbf{1} \left(\alpha_i \geq \frac{1}{2(|\mathcal{N}| - f)} \right) \geq |\mathcal{N}| - f \quad \} \end{aligned} \quad (91)$$

³ In contrast to Algorithms 1, 2 and 3 in [19], the adopted default tuple in Algorithm 3 here is not necessarily known to all agents. In addition, the default tuple may vary across iterations.

Each $p(x) \in \tilde{\mathcal{C}}$ is called a valid function. Note that the function $\frac{1}{|\mathcal{N}|} \sum_{i \in \mathcal{N}} h_i(x) \in \tilde{\mathcal{C}}$ since $n \geq 3f + 1$ and $|\mathcal{N}| \geq 2f + 1$. For ease of future reference, we let $\tilde{p}(x) = \frac{1}{|\mathcal{N}|} \sum_{i \in \mathcal{N}} h_i(x)$. Define $\tilde{Y} \triangleq \cup_{p(x) \in \tilde{\mathcal{C}}} \text{argmin } p(x)$.

Lemma 10. [19] \tilde{Y} is a convex set.

Lemma 11. \tilde{Y} is a closed set.

Lemma 11 is proved in Appendix G. By Lemma 11, Definition 1 is well-defined over \tilde{Y} .

4.1 Update Dynamic – Matrix Representation

Definition 2. [23] For a given graph $G(\mathcal{V}, \mathcal{E})$, a reduced graph \mathcal{H} is a subgraph of $G(\mathcal{V}, \mathcal{E})$ obtained by (i) removing all the faulty agents from \mathcal{V} along with their edges; (ii) removing any additional up to f incoming edges at each non-faulty agent.

Let us denote the collection of all the reduced graphs for a given $G(\mathcal{V}, \mathcal{E})$ by $\mathcal{R}_{\mathcal{F}}$. Thus, $\mathcal{V} - \mathcal{F}$ is the set of agents in each element in $\mathcal{R}_{\mathcal{F}}$. Let $\tau = |\mathcal{R}_{\mathcal{F}}|$. It is easy to see that τ depends on \mathcal{F} , and it is finite.

Without loss of generality, assume agents indexed from 1 through $n - \phi$ are non-faulty, and agents indexed from $n - \phi + 1$ to n are faulty. Let $\mathbf{x}[t - 1] \in \mathbb{R}^{n - \phi}$ be a real vector of the local estimates at the beginning of iteration t with $\mathbf{x}_j[t - 1] = x_j[t - 1]$ being the local estimate of agent $j \in \mathcal{N}$, and let $\tilde{\mathbf{g}}[t - 1] \in \mathbb{R}^{n - \phi}$ be a vector of the local gradients at iteration t with $\tilde{\mathbf{g}}_j[t - 1] = \tilde{g}_j[t - 1], j \in \mathcal{N}$. Since the underlying communication network is a complete graph with $n \geq 3f + 1$, as shown in [22], the update of $\mathbf{x} \in \mathbb{R}^{n - \phi}$ in each iteration can be written compactly in a matrix form.

$$\mathbf{x}[t] = \mathbf{M}[t - 1]\mathbf{x}[t - 1] - \lambda[t - 1]\tilde{\mathbf{g}}[t - 1]. \quad (92)$$

The construction of $\mathbf{M}[t]$ and relevant properties are given in [22]. Let $\mathcal{H} \in \mathcal{R}_{\mathcal{F}}$ be a reduced graph of the given communication graph, with \mathbf{H} as the adjacency matrix. It is shown in [22] that in every iteration t , and for every $\mathbf{M}[t]$, there exists a reduced graph $\mathcal{H}[t] \in \mathcal{R}_{\mathcal{F}}$ with adjacency matrix $\mathbf{H}[t]$ such that

$$\mathbf{M}[t] \geq \beta \mathbf{H}[t], \quad (93)$$

where $0 < \beta < 1$ is a constant. The definition of β can be found in [22].

Equation (92) can be further expanded out as

$$\begin{aligned} \mathbf{x}[t] &= \mathbf{M}[t - 1]\mathbf{x}[t - 1] - \lambda[t - 1]\tilde{\mathbf{g}}[t - 1] \\ &= \mathbf{M}[t - 1](\mathbf{M}[t - 2]\mathbf{x}[t - 2] - \lambda[t - 2]\tilde{\mathbf{g}}[t - 2]) - \lambda[t - 1]\tilde{\mathbf{g}}[t - 1] \\ &= \mathbf{M}[t - 1]\mathbf{M}[t - 2]\mathbf{x}[t - 2] - \lambda[t - 2]\mathbf{M}[t - 1]\tilde{\mathbf{g}}[t - 2] - \lambda[t - 1]\tilde{\mathbf{g}}[t - 1] \\ &= \dots \\ &= (\mathbf{M}[t - 1]\mathbf{M}[t - 2] \cdots \mathbf{M}[0]\mathbf{x}[0]) - \lambda[0](\mathbf{M}[t - 1]\mathbf{M}[t - 2] \cdots \mathbf{M}[1]\tilde{\mathbf{g}}[0]) - \dots - \\ &\quad - \lambda[t - 1]\tilde{\mathbf{g}}[t - 1] \\ &= \Phi(t - 1, 0)\mathbf{x}[0] - \sum_{r=0}^{t-1} \lambda[r]\Phi(t - 1, r + 1)\tilde{\mathbf{g}}[r], \end{aligned} \quad (94)$$

where $\Phi(t - 1, r) = \mathbf{M}[t - 1]\mathbf{M}[t - 2] \cdots \mathbf{M}[r]$ is a backward product, and by convention, $\Phi(t - 1, t - 1) = \mathbf{M}[t - 1]$ and $\Phi(t - 1, t) = \mathbf{I}$.

4.2 Correctness of Algorithm 3

Using coefficients of ergodicity theorem, it is showed in [22] that $\Phi(t, r)$ is weak-ergodic [22], and that the rate of the convergence is exponential [2], as formally stated in Theorem 3. Recall that $\tau = |R_{\mathcal{F}}|$, $n - \phi$ is the total number of non-faulty agents, and $0 < \beta < 1$ is a constant for which (93) holds.

Theorem 3. [2] Let $\nu = \tau(n - \phi)$ and $\gamma = 1 - \beta^\nu$. For any sequence $\Phi(t, r)$,

$$|\Phi_{ik}(t, r) - \Phi_{jk}(t, r)| \leq \gamma^{\lceil \frac{t-r+1}{\nu} \rceil}, \quad (95)$$

for all $t \geq r$.

Lemma 12. For all $i, j \in \mathcal{N}$ and for each $t \geq 1$,

$$|x_i[t] - x_j[t]| \leq (n - \phi) \max\{|u|, |U|\} \gamma^{\lceil \frac{t}{\nu} \rceil} + L \sum_{r=0}^{t-1} \lambda[r] (n - \phi) \gamma^{\lceil \frac{t-1-r}{\nu} \rceil},$$

and for all $i, j \in \mathcal{N}$ and for $t = 0$,

$$|x_i[0] - x_j[0]| \leq U - u.$$

The proof of Lemma 12 can be found in Appendix D.

Corollary 3. For $i, j \in \mathcal{N}$,

$$\lim_{t \rightarrow \infty} |x_i[t] - x_j[t]| = 0.$$

We present the proof of Corollary 3 in Appendix E.

Let $M[t] = \max_{i \in \mathcal{N}} x_i[t]$ and $m[t] = \min_{i \in \mathcal{N}} x_i[t]$. The following lemma holds.

Lemma 13. Under Algorithm 3, the following holds.

$$\sum_{t=0}^{\infty} \lambda[t] (M[t] - m[t]) < \infty.$$

The proof of Lemma 13 is similar to the proof of Lemma 8. For completeness, we present the proof in Appendix F.

Proposition 2. Let $a, b, c, d \in \mathbb{R}$ such that $b < a, b \leq c \leq \frac{1}{2}(a + b), \frac{1}{2}(a + b) < a \leq d$, and there exists $0 \leq \xi \leq 1$, for which $\frac{1}{2}(a + b) = \xi d + (1 - \xi)c$ holds. Then

$$\frac{1}{2} \leq \xi \leq 1.$$

Proof. Suppose, on the contrary, that $0 \leq \xi < \frac{1}{2}$. Since $c \geq b$ and $d \geq a$, we have

$$\frac{1}{2}(c + d) \geq \frac{1}{2}(a + b) \quad (96)$$

On the other hand, by the assumptions that $a > b$, and that $\frac{1}{2}(a + b) \geq c$, it holds that

$$d \geq a > \frac{1}{2}(a + b) \geq c, \quad (97)$$

i.e., $d > c$. Then

$$\begin{aligned}
\frac{1}{2}(c+d) &= \frac{1}{2}c + \frac{1}{2}d \\
&= \xi c + \left(\frac{1}{2} - \xi\right)c + \frac{1}{2}d \\
&< \xi c + \left(\frac{1}{2} - \xi\right)d + \frac{1}{2}d \quad \text{by (97)} \\
&= \xi c + \left(\frac{1}{2} - \xi + \frac{1}{2}\right)d \\
&= \xi c + (1 - \xi)d \\
&= \frac{1}{2}(a+b), \tag{98}
\end{aligned}$$

i.e., $\frac{1}{2}(c+d) < \frac{1}{2}(a+b)$. The relations in (96) and (98) contradict each other. Thus, the assumption that $0 \leq \xi < \frac{1}{2}$ does not hold, i.e., $\frac{1}{2} \leq \xi \leq 1$, proving the proposition. \square

Lemma 14. *For each non-faulty agent $j \in \mathcal{N}$ and each iteration $t \geq 1$, there exists a valid function $p(x) = \sum_{i \in \mathcal{N}} \alpha_i h_i(x) \in \mathcal{C}$ such that*

$$\tilde{g}_j[t-1] = \sum_{i \in \mathcal{N}} \alpha_i h'_i(x_i[t-1]).$$

Proof. Recall that $\mathcal{R}_j^2[t-1]$ denotes the set of agents from whom the remaining $n - 2f$ gradient values (second entries of the tuples) were received in iteration t , and let us denote by $\mathcal{L}_j[t-1]$ and $\mathcal{S}_j[t-1]$ the set of agents from whom the largest f gradient values and the smallest f gradient values were received in iteration t .

Let $i^*, j^* \in \mathcal{R}_j^2[t-1]$ such that $g_{i^*}[t-1] = \check{g}_j[t-1]$ and $g_{j^*}[t-1] = \hat{g}_j[t-1]$. Recall that $|\mathcal{F}| = \phi$. Let $\mathcal{L}_j^*[t-1] \subseteq \mathcal{L}_j[t-1] - \mathcal{F}$ and $\mathcal{S}_j^*[t-1] \subseteq \mathcal{S}_j[t-1] - \mathcal{F}$ such that

$$|\mathcal{L}_j^*[t-1]| = f - \phi + |\mathcal{R}_j^2[t-1] \cap \mathcal{F}|,$$

and

$$|\mathcal{S}_j^*[t-1]| = f - \phi + |\mathcal{R}_j^2[t-1] \cap \mathcal{F}|.$$

We consider two cases: (i) $\hat{g}_j[t-1] > \check{g}_j[t-1]$ and (ii) $\hat{g}_j[t-1] = \check{g}_j[t-1]$, separately.

Case (i): $\hat{g}_j[t-1] > \check{g}_j[t-1]$. By definition of $\mathcal{L}_j^*[t-1]$ and $\mathcal{S}_j^*[t-1]$, we have

$$\frac{1}{f - \phi + |\mathcal{R}_j^2[t-1] \cap \mathcal{F}|} \sum_{i \in \mathcal{S}_j^*[t-1]} g_i[t-1] \leq \tilde{g}_j[t-1] \leq \frac{1}{f - \phi + |\mathcal{R}_j^2[t-1] \cap \mathcal{F}|} \sum_{i \in \mathcal{L}_j^*[t-1]} g_i[t-1]. \tag{99}$$

Thus, there exists $0 \leq \xi \leq 1$ such that

$$\begin{aligned}
\tilde{g}_j[t-1] &= \xi \left(\frac{1}{f - \phi + |\mathcal{R}_j^2[t-1] \cap \mathcal{F}|} \sum_{i \in \mathcal{S}_j^*[t-1]} g_i[t-1] \right) + (1 - \xi) \left(\frac{1}{f - \phi + |\mathcal{R}_j^2[t-1] \cap \mathcal{F}|} \sum_{i \in \mathcal{L}_j^*[t-1]} g_i[t-1] \right) \\
&= \frac{\xi}{f - \phi + |\mathcal{R}_j^2[t-1] \cap \mathcal{F}|} \sum_{i \in \mathcal{S}_j^*[t-1]} g_i[t-1] + \frac{1 - \xi}{f - \phi + |\mathcal{R}_j^2[t-1] \cap \mathcal{F}|} \sum_{i \in \mathcal{L}_j^*[t-1]} g_i[t-1]. \tag{100}
\end{aligned}$$

By symmetry, WLOG, assume $\xi \geq \frac{1}{2}$.

Let $k \in \mathcal{R}_j^2[t-1] - \mathcal{F}$. By symmetry, WLOG, assume $g_k[t-1] \leq \tilde{g}_j[t-1]$. Since $|\mathcal{L}_j[t-1] \cup \{j^*\}| = f+1$, there exists a non-faulty agent $j'_k \in \mathcal{L}_j[t-1] \cup \{j^*\}$. Thus, $g_{j'_k}[t-1] \geq \hat{g}_j[t-1] > \tilde{g}_j[t-1]$, and there exists $0 \leq \xi_k \leq 1$ such that

$$\frac{1}{2} (\hat{g}_j[t-1] + \check{g}_j[t-1]) = \tilde{g}_j[t-1] = \xi_k g_k[t-1] + (1 - \xi_k) g_{j'_k}[t-1]. \quad (101)$$

Let $a = \hat{g}_j[t-1]$, $b = \check{g}_j[t-1]$, $c = g_k[t-1]$, and $d = g_{j'_k}[t-1]$. By Proposition 2, we know that $\frac{1}{2} \leq \xi_k \leq 1$.

Since $|\mathcal{N}| - f = n - \phi - f = n - 2f + f - \phi = |\mathcal{R}_j^2[t-1]| + f - \phi = |\mathcal{R}_j^2[t-1] - \mathcal{F}| + |\mathcal{R}_j^2[t-1] \cap \mathcal{F}| + f - \phi$, we get

$$\begin{aligned} \tilde{g}_j[t-1] &= \frac{|\mathcal{N}| - f}{|\mathcal{N}| - f} \tilde{g}_j[t-1] \\ &= \frac{|\mathcal{R}_j^2[t-1] - \mathcal{F}|}{|\mathcal{N}| - f} \tilde{g}_j[t-1] + \frac{f - \phi + |\mathcal{R}_j^2[t-1] \cap \mathcal{F}|}{|\mathcal{N}| - f} \tilde{g}_j[t-1] \\ &= \frac{1}{|\mathcal{N}| - f} \left(\sum_{k \in \mathcal{R}_j^2[t-1] - \mathcal{F}} \tilde{g}_j[t-1] \right) + \frac{f - \phi + |\mathcal{R}_j^2[t-1] \cap \mathcal{F}|}{|\mathcal{N}| - f} \tilde{g}_j[t-1] \\ &= \frac{1}{|\mathcal{N}| - f} \sum_{k \in \mathcal{R}_j^2[t-1] - \mathcal{F}} (\xi_k g_k[t-1] + (1 - \xi_k) g_{j'_k}[t-1]) \\ &\quad + \frac{\xi}{|\mathcal{N}| - f} \sum_{i \in \mathcal{S}_j^*[t-1]} g_i[t-1] + \frac{1 - \xi}{|\mathcal{N}| - f} \sum_{i \in \mathcal{L}_j^*[t-1]} g_i[t-1] \quad \text{by (100) and (101)} \\ &= \frac{1}{|\mathcal{N}| - f} \sum_{k \in \mathcal{R}_j^2[t-1] - \mathcal{F}} (\xi_k h'_k(x_k[t-1]) + (1 - \xi_k) h'_{j'_k}(x_{j'_k}[t-1])) \\ &\quad + \frac{\xi}{|\mathcal{N}| - f} \sum_{i \in \mathcal{S}_j^*[t-1]} h'_i(x_i[t-1]) + \frac{1 - \xi}{|\mathcal{N}| - f} \sum_{i \in \mathcal{L}_j^*[t-1]} h'_i(x_i[t-1]). \end{aligned}$$

Define $q(x)$ as follows.

$$\begin{aligned} q(x) &= \frac{1}{|\mathcal{N}| - f} \sum_{k \in \mathcal{R}_j^2[t-1] - \mathcal{F}} (\xi_k h_k(x) + (1 - \xi_k) h_{j'_k}(x)) \\ &\quad + \frac{\xi}{|\mathcal{N}| - f} \sum_{i \in \mathcal{S}_j^*[t-1]} h_i(x) + \frac{1 - \xi}{|\mathcal{N}| - f} \sum_{i \in \mathcal{L}_j^*[t-1]} h_i(x). \end{aligned} \quad (102)$$

In (102), for each $k \in \mathcal{R}_j^2[t-1] - \mathcal{F}$, it holds that $\frac{\xi_k}{|\mathcal{N}| - f} \geq \frac{1}{2(|\mathcal{N}| - f)}$. For each $i \in \mathcal{S}_j^*[t-1]$, it holds that $\frac{\xi}{|\mathcal{N}| - f} \geq \frac{1}{2(|\mathcal{N}| - f)}$. In addition, we have

$$\begin{aligned} |(\mathcal{R}_j^2[t-1] - \mathcal{F}) \cup \mathcal{S}_j^*[t-1]| &= |\mathcal{R}_j^2[t-1] - \mathcal{F}| + |\mathcal{S}_j^*[t-1]| \\ &= |\mathcal{R}_j^2[t-1]| - |\mathcal{R}_j^2[t-1] \cap \mathcal{F}| + |\mathcal{S}_j^*[t-1]| \\ &= n - 2f - |\mathcal{R}_j^2[t-1] \cap \mathcal{F}| + f - \phi + |\mathcal{R}_j^2[t-1] \cap \mathcal{F}| \\ &= n - \phi - f = |\mathcal{N}| - f. \end{aligned}$$

Thus, in (102), at least $|\mathcal{N}| - f$ non-faulty agents corresponding to agents $k \in (\mathcal{R}_j^2[t-1] - \mathcal{F}) \cup \mathcal{S}_j^*[t-1]$ are assigned with weights lower bounded by $\frac{1}{2(|\mathcal{N}| - f)}$.

Case (ii): $\hat{g}_j[t-1] = \check{g}_j[t-1]$. Let $k \in \mathcal{R}_j^2[t-1] - \mathcal{F}$. Since $\hat{g}_j[t-1] \geq g_k[t-1] \geq \check{g}_j[t-1]$ and $\hat{g}_j[t-1] = \check{g}_j[t-1]$, it holds that $\hat{g}_j[t-1] = g_k[t-1] = \check{g}_j[t-1]$. Consequently, we have

$$\tilde{g}_j[t-1] = \frac{1}{2} (\hat{g}_j[t-1] + \check{g}_j[t-1]) = g_k[t-1].$$

So we can rewrite $\tilde{g}_j[t-1]$ as follows.

$$\begin{aligned} \tilde{g}_j[t-1] &= \frac{|\mathcal{N}| - f}{|\mathcal{N}| - f} \tilde{g}_j[t-1] \\ &= \frac{1}{|\mathcal{N}| - f} \left(\sum_{k \in \mathcal{R}_j^2[t-1] - \mathcal{F}} \tilde{g}_j[t-1] \right) + \frac{f - \phi + |\mathcal{R}_j^2[t-1] \cap \mathcal{F}|}{|\mathcal{N}| - f} \tilde{g}_j[t-1] \\ &= \frac{1}{|\mathcal{N}| - f} \sum_{k \in \mathcal{R}_j^2[t-1] - \mathcal{F}} g_k[t-1] + \frac{\xi}{|\mathcal{N}| - f} \sum_{i \in \mathcal{S}_j^*[t-1]} g_i[t-1] + \frac{1 - \xi}{|\mathcal{N}| - f} \sum_{i \in \mathcal{L}_j^*[t-1]} g_i[t-1] \\ &= \frac{1}{|\mathcal{N}| - f} \sum_{k \in \mathcal{R}_j^2[t-1] - \mathcal{F}} h'_k(x_k[t-1]) \\ &\quad + \frac{\xi}{|\mathcal{N}| - f} \sum_{i \in \mathcal{S}_j^*[t-1]} h'_i(x_i[t-1]) + \frac{1 - \xi}{|\mathcal{N}| - f} \sum_{i \in \mathcal{L}_j^*[t-1]} h'_i(x_i[t-1]). \end{aligned}$$

Define $q(x)$ as follows.

$$q(x) = \frac{1}{|\mathcal{N}| - f} \sum_{k \in \mathcal{R}_j^2[t-1] - \mathcal{F}} h_k(x) + \frac{\xi}{|\mathcal{N}| - f} \sum_{i \in \mathcal{S}_j^*[t-1]} h_i(x) + \frac{1 - \xi}{|\mathcal{N}| - f} \sum_{i \in \mathcal{L}_j^*[t-1]} h_i(x). \quad (103)$$

In (103), for each $k \in \mathcal{R}_j^2[t-1] - \mathcal{F}$, it holds that $\frac{1}{|\mathcal{N}| - f} \geq \frac{1}{2(|\mathcal{N}| - f)}$. For each $i \in \mathcal{S}_j^*[t-1]$, it holds that $\frac{\xi}{|\mathcal{N}| - f} \geq \frac{1}{2(|\mathcal{N}| - f)}$. In addition, we have

$$|(\mathcal{R}_j^2[t-1] - \mathcal{F}) \cup \mathcal{S}_j^*[t-1]| = |\mathcal{N}| - f.$$

Thus, in (103), at least $|\mathcal{N}| - f$ non-faulty agents corresponding to $(\mathcal{R}_j^2[t-1] - \mathcal{F}) \cup \mathcal{S}_j^*[t-1]$ are assigned with weights lower bounded by $\frac{1}{2(|\mathcal{N}| - f)}$.

Case (i) and Case (ii) together prove the lemma. \square

Proposition 3. For each non-faulty agent $j \in \mathcal{N}$ and each $t \geq 1$, there exists a set of convex coefficients β_i 's over non-faulty agents, i.e., $\beta_i \geq 0$ for each $i \in \mathcal{N}$ and $\sum_{i \in \mathcal{N}} \beta_i = 1$, such that the following holds

$$\frac{1}{n - 2f} \sum_{i \in \mathcal{R}_j^1[t-1]} w_i[t-1] = \sum_{i \in \mathcal{N}} \beta_i x_i[t-1].$$

Proof. Note that $\mathcal{R}_j^1[t-1] = (\mathcal{R}_j^1[t-1] - \mathcal{F}) \cup (\mathcal{R}_j^1[t-1] \cap \mathcal{F})$. We consider two cases: (i) $\mathcal{R}_j^1[t-1] \cap \mathcal{F} = \emptyset$ and (ii) $\mathcal{R}_j^1[t-1] \cap \mathcal{F} \neq \emptyset$, separately.

Case (i): $\mathcal{R}_j^1[t-1] \cap \mathcal{F} = \emptyset$. When $\mathcal{R}_j^1[t-1] \cap \mathcal{F} = \emptyset$, every agent in $\mathcal{R}_j^1[t-1]$ is non-faulty, i.e., $\mathcal{R}_j^1[t-1] \subseteq \mathcal{N}$. Then we get

$$\frac{1}{n-2f} \sum_{i \in \mathcal{R}_j^1[t-1]} w_i[t-1] = \frac{1}{n-2f} \sum_{i \in \mathcal{R}_j^1[t-1]} x_i[t-1] \quad \text{since } w_i[t-1] = x_i[t-1] \text{ for each } i \in \mathcal{N} \quad (104)$$

Let $\beta_i = \frac{1}{n-2f}$ for each $i \in \mathcal{R}_j^1[t-1] \subseteq \mathcal{N}$, and $\beta_i = 0$ for each $i \in \mathcal{N} - \mathcal{R}_j^1[t-1]$. The obtained β_i 's is a valid collection of convex coefficients, since $\beta_i = \frac{1}{n-2f} \geq 0$ for each $i \in \mathcal{N}$, and

$$\sum_{i \in \mathcal{N}} \beta_i = \sum_{i \in \mathcal{R}_j^1[t-1]} \beta_i = \sum_{i \in \mathcal{R}_j^1[t-1]} \frac{1}{n-2f} = \frac{1}{n-2f} |\mathcal{R}_j^1[t-1]| = \frac{1}{n-2f} (n-2f) = 1.$$

Case (ii): $\mathcal{R}_j^1[t-1] \cap \mathcal{F} \neq \emptyset$. Let $\mathcal{L}_j[t-1]$ be the set of the identifiers of the f agents from whom the f largest first entries ($w_i[t-1]$'s) are received, and let $\mathcal{S}_j[t-1]$ be the set of the identifiers of the f agents from whom the f smallest first entries ($w_i[t-1]$'s) are received. Since $\mathcal{R}_j^1[t-1] \cap \mathcal{F} \neq \emptyset$, it holds that $\mathcal{L}_j[t-1] \cap \mathcal{N} \neq \emptyset$ and $\mathcal{S}_j[t-1] \cap \mathcal{N} \neq \emptyset$. Let l and s be two non-faulty agents such that $l \in \mathcal{L}_j[t-1] \cap \mathcal{N}$ and $s \in \mathcal{S}_j[t-1] \cap \mathcal{N}$. By definition of $\mathcal{R}_j^1[t-1]$, for each $k \in \mathcal{R}_j^1[t-1] \cap \mathcal{F}$, we have

$$x_s[t-1] = w_s[t-1] \leq w_k[t-1] \leq w_l[t-1] = x_l[t-1]. \quad (105)$$

Then,

$$|\mathcal{R}_j^1[t-1] \cap \mathcal{F}| x_s[t-1] \leq \sum_{k \in \mathcal{R}_j^1[t-1] \cap \mathcal{F}} w_k[t-1] \leq |\mathcal{R}_j^1[t-1] \cap \mathcal{F}| x_l[t-1].$$

Thus, there exists $0 \leq \zeta \leq 1$ such that

$$\sum_{k \in \mathcal{R}_j^1[t-1] \cap \mathcal{F}} w_k[t-1] = \zeta (|\mathcal{R}_j^1[t-1] \cap \mathcal{F}| x_s[t-1]) + (1-\zeta) (|\mathcal{R}_j^1[t-1] \cap \mathcal{F}| x_l[t-1]). \quad (106)$$

Thus,

$$\begin{aligned} \frac{1}{n-2f} \sum_{i \in \mathcal{R}_j^1[t-1]} w_i[t-1] &= \frac{1}{n-2f} \left(\sum_{i \in \mathcal{R}_j^1[t-1] - \mathcal{F}} w_i[t-1] + \sum_{i \in \mathcal{R}_j^1[t-1] \cap \mathcal{F}} w_i[t-1] \right) \\ &= \frac{1}{n-2f} \left(\sum_{i \in \mathcal{R}_j^1[t-1] - \mathcal{F}} x_i[t-1] + \sum_{i \in \mathcal{R}_j^1[t-1] \cap \mathcal{F}} w_i[t-1] \right) \quad \text{since } w_i[t-1] = x_i[t-1] \text{ for each } i \in \mathcal{N} \\ &= \frac{1}{n-2f} \left(\sum_{i \in \mathcal{R}_j^1[t-1] - \mathcal{F}} x_i[t-1] + \zeta (|\mathcal{R}_j^1[t-1] \cap \mathcal{F}| x_s[t-1]) + (1-\zeta) (|\mathcal{R}_j^1[t-1] \cap \mathcal{F}| x_l[t-1]) \right) \quad \text{by (106)} \\ &= \frac{1}{n-2f} \sum_{i \in \mathcal{R}_j^1[t-1] - \mathcal{F}} x_i[t-1] + \frac{\zeta |\mathcal{R}_j^1[t-1] \cap \mathcal{F}|}{n-2f} x_s[t-1] + \frac{(1-\zeta) |\mathcal{R}_j^1[t-1] \cap \mathcal{F}|}{n-2f} x_l[t-1]. \end{aligned} \quad (107)$$

Let $\beta_s = \frac{\zeta|\mathcal{R}_j^1[t-1] \cap \mathcal{F}|}{n-2f}$, $\beta_l = \frac{(1-\zeta)|\mathcal{R}_j^1[t-1] \cap \mathcal{F}|}{n-2f}$, let $\beta_i = \frac{1}{n-2f}$ for each $i \in \mathcal{R}_j^1[t-1] - \mathcal{F}$, and let $\beta_i = 0$ for all other non-faulty agents. The obtained β_i 's is a valid collection of convex coefficients since

$$\begin{aligned} \beta_s + \beta_l + \sum_{i \in \mathcal{R}_j^1[t-1] - \mathcal{F}} \beta_i &= \frac{\zeta|\mathcal{R}_j^1[t-1] \cap \mathcal{F}|}{n-2f} + \frac{(1-\zeta)|\mathcal{R}_j^1[t-1] \cap \mathcal{F}|}{n-2f} + \sum_{i \in \mathcal{R}_j^1[t-1] - \mathcal{F}} \frac{1}{n-2f} \\ &= \frac{|\mathcal{R}_j^1[t-1] \cap \mathcal{F}|}{n-2f} + \frac{|\mathcal{R}_j^1[t-1] - \mathcal{F}|}{n-2f} \\ &= \frac{|\mathcal{R}_j^1[t-1]|}{n-2f} = \frac{n-2f}{n-2f} = 1. \end{aligned}$$

Case (i) and case (ii) together prove the proposition. \square

We define $z[t]$ and x_{j_t} similar to that for Algorithm 1. In particular, let $\{z[t]\}_{t=0}^\infty$ be a sequence of estimates such that

$$z[t] = x_{j_t}[t], \quad \text{where } j_t \in \operatorname{argmax}_{j \in \mathcal{N}[t]} \operatorname{Dist}(x_j[t], Y). \quad (108)$$

From the definition, there is a sequence of agents $\{j_t\}_{t=0}^\infty$ associated with the sequence $\{z[t]\}_{t=0}^\infty$.

Theorem 4. *The sequence $\{\operatorname{Dist}(z[t], Y)\}_{t=0}^\infty$ converges and*

$$\lim_{t \rightarrow \infty} \operatorname{Dist}(z[t], Y) = 0.$$

Proof.

$$\begin{aligned} \operatorname{Dist}(z[t+1], Y) &= \operatorname{Dist}(x_{j_{t+1}}[t], Y) \quad \text{by (108)} \\ &= \operatorname{Dist}\left(\frac{1}{n-2f} \sum_{i \in \mathcal{R}_{j_{t+1}}^1[t]} w_i[t] - \lambda[t] \tilde{g}_{j_{t+1}}[t], Y\right) \quad \text{by (90)} \\ &= \operatorname{Dist}\left(\sum_{i \in \mathcal{N}} \beta_i x_i[t] - \lambda[t] \tilde{g}_{j_{t+1}}[t], Y\right) \quad \text{by Proposition 3} \\ &= \operatorname{Dist}\left(\sum_{i \in \mathcal{N}} \beta_i (x_i[t] - \lambda[t] \tilde{g}_{j_{t+1}}[t]), Y\right) \quad \text{since } \sum_{i \in \mathcal{N}} \beta_i = 1 \\ &\leq \sum_{i \in \mathcal{N}} \beta_i \operatorname{Dist}(x_i[t] - \lambda[t] \tilde{g}_{j_{t+1}}[t], Y) \quad \text{by convexity of } \operatorname{Dist}(\cdot, Y) \\ &\leq \max_{i \in \mathcal{N}} \operatorname{Dist}(x_i[t] - \lambda[t] \tilde{g}_{j_{t+1}}[t], Y). \end{aligned} \quad (109)$$

By Lemma 14, there exists a valid function $p_t(\cdot) = \sum_{q \in \mathcal{N}} \alpha_q h_q(\cdot) \in \mathcal{C}$ such that

$$\tilde{g}_{j_{t+1}}[t] = \sum_{q \in \mathcal{N}} \alpha_q h'_q(x_q[t]). \quad (110)$$

In addition, let

$$j'_{t+1} \in \operatorname{argmax}_{i \in \mathcal{N}} \operatorname{Dist}(x_i[t] - \lambda[t] \tilde{g}_{j_{t+1}}[t], Y).$$

We get

$$\begin{aligned}
\text{Dist}(z[t+1], Y) &\leq \max_{i \in \mathcal{N}} \text{Dist}(x_i[t] - \lambda[t] \tilde{g}_{j_{t+1}}[t], Y) \quad \text{by (109)} \\
&= \text{Dist}(x_{j'_{t+1}}[t] - \lambda[t] \tilde{g}_{j_{t+1}}[t], Y) \\
&= \text{Dist}\left(x_{j'_{t+1}}[t] - \lambda[t] \sum_{q \in \mathcal{N}} \alpha_q h'_q(x_q[t]), Y\right) \quad \text{by Lemma 14} \\
&= \inf_{y \in Y} \left| x_{j'_{t+1}}[t] - \lambda[t] \sum_{q \in \mathcal{N}} \alpha_q h'_q(x_q[t]) - y \right| \\
&\leq \inf_{y \in Y} \left| x_{j'_{t+1}}[t] - \lambda[t] p'_{t+1}(x_{j'_{t+1}}[t]) - y \right| + \lambda[t] L(M[t] - m[t]). \quad (111)
\end{aligned}$$

where p_t is defined in (110). Note that for each $t \geq 0$, there exists a non-faulty agent j'_t such that (111) holds, and there exists a sequence of agents $\{j'_t\}_{t=0}^\infty$. Let $\{x[t]\}_{t=0}^\infty$ be a sequence of estimates such that $x[t] = x_{j'_t}[t]$. Let $\{g[t]\}_{t=0}^\infty$ be a sequence of gradients such that $g[t] = p'_t(x_{j'_t}[t])$.

The remaining of the proof is identical to the proof of Theorem 2. □

5 Discussion and Conclusion

So far, a synchronous system is considered. In an asynchronous system, when there are up to f crash faults, Problem 1 is not solvable, since it is possible that every agent in the system is non-faulty, but f agents are slow. In this case, the system will mistakenly “treat” the slow agents as crashed agents. Consequently, the weights of the slow agents may be strictly smaller than the other agents. Despite the impossibility of solving Problem 1 in asynchronous system, nevertheless, Problem 2 can be solved with $\beta \geq \frac{1}{n}$ and $\gamma \geq |\mathcal{N}| - f$. In particular, Algorithm 2 can be easily adapted for asynchronous system by modifying the receiving step (step 2). For completeness, we list out the algorithm for crash faults.

Algorithm 4 (crash faults) for agent j for iteration $t \geq 1$:

Step 1: Compute $h'_j(x_j[t-1])$ – the gradient of local function $h_j(\cdot)$ at point $x_j[t-1]$, and send the triple $(x_j[t-1], h'_j(x_j[t-1]), t)$ to all the agents (including agent j itself).

Step 2: Upon receiving $(x_i[t-1], h'_i(x_i[t-1]), t)$ from $n - f$ non-faulty agents (including agent j itself) – these received tuples form a multiset $\mathcal{R}_j[t-1]$, update x_j as

$$x_j[t] = \frac{1}{|\mathcal{R}_j[t-1]|} \left(\sum_{i \in \mathcal{R}_j[t-1]} (x_i[t-1] - \lambda[t-1] h'_i(x_i[t-1])) \right). \quad (112)$$

Note that $|\mathcal{R}_j[t-1]| = n - f$. Since at most f agents may crash, agent j can receive messages from at least $n - f$ agents in step 2. Thus, Algorithm 3 will always proceed to the next iteration. We are able to show the following theorem.

Theorem 5. *Algorithm 4 solves Problem 2 with $\beta = \frac{1}{n}$ and $\gamma = n - f$.*

The collection of valid function is defined as follows.

$$\mathcal{C} \triangleq \left\{ p(x) : p(x) = \sum_{i \in \mathcal{V}} \alpha_i h_i(x), \forall i \in \mathcal{V}, \alpha_i \geq 0, \sum_{i \in \mathcal{N}} \alpha_i = 1, \text{ and } \sum_{i \in \mathcal{V}} \mathbf{1} \left(\alpha_i \geq \frac{1}{n} \right) \geq n - f \right\}$$

The proof of Theorem 5 is similar to the proof of Theorem 2.

In an asynchronous system, when there are up to f Byzantine faults, simple iterative algorithms like Algorithm 3 may not exist, observing that it is impossible to achieve Byzantine consensus with single round of message exchange with only $n = 3f + 1$ agents. In contrast, when the algorithm introduced in [1] is used as a communication mechanism in each iteration, we believe that Algorithm 3 can be modified such that it can solve Problem 2 with $\beta \geq \frac{1}{2(|\mathcal{N}| - f)}$ and $\gamma \geq |\mathcal{N}| - 2f$. There may be a tradeoff between the system size n and the communication load in each iteration. We leave this problem for future exploration.

Note that the definition of admissibility of the local functions in this report is slightly different from that in [19]. Comparing to [19], stronger assumptions are used in proving the correctness of the three iterative algorithms developed in this work. In particular, we require that the local functions have to have L -Lipschitz derivatives. Whether such assumptions are necessary or not is still open, and we leave this for future exploration as well.

References

1. Ittai Abraham, Yonatan Amit, and Danny Dolev. Optimal resilience asynchronous approximate agreement. In *Principles of Distributed Systems*, pages 229–239. Springer, 2005.
2. Jac.M Anthonisse and Henk Tijms. Exponential convergence of products of stochastic matrices. *Journal of Mathematical Analysis and Applications*, 59(2):360 – 364, 1977.
3. Dimitri P Bertsekas and John N Tsitsiklis. *Parallel and distributed computation: numerical methods*. Prentice-Hall, Inc., 1989.
4. Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.*, 3(1):1–122, January 2011.
5. Soma Chaudhuri. More choices allow more faults: Set consensus problems in totally asynchronous systems. *Information and Computation*, 105:132–158, 1992.
6. Danny Dolev, Nancy A. Lynch, Shlomit S. Pinter, Eugene W. Stark, and William E. Weihl. Reaching approximate agreement in the presence of faults. *J. ACM*, 33(3):499–516, May 1986.
7. J.C. Duchi, A. Agarwal, and M.J. Wainwright. Dual averaging for distributed optimization: Convergence analysis and network scaling. *Automatic Control, IEEE Transactions on*, 57(3):592–606, March 2012.
8. Alan David Fekete. Asymptotically optimal algorithms for approximate agreement. *Distributed Computing*, 4(1):9–29, 1990.
9. Michael J. Fischer, Nancy A. Lynch, and Michael Merritt. Easy impossibility proofs for distributed consensus problems. In *Proceedings of the fourth annual ACM symposium on Principles of distributed computing*, PODC ’85, pages 59–70, New York, NY, USA, 1985. ACM.
10. Roy Friedman, Achour Mostefaoui, Sergio Rajsbaum, and Michel Raynal. Asynchronous agreement and its relation with error-correcting codes. *Computers, IEEE Transactions on*, 56(7):865–875, 2007.
11. Bhavya Kailkhura, Swastik Brahma, and Pramod K Varshney. Consensus based detection in the presence of data falsification attacks. *arXiv preprint arXiv:1504.03413*, 2015.
12. Heath J. LeBlanc, Haotian Zhang, Shreyas Sundaram, and Xenofon Koutsoukos. Consensus of multi-agent networks in the presence of adversaries using only local information. In *Proceedings of the 1st International Conference on High Confidence Networked Systems*, HiCoNS ’12, pages 1–10, New York, NY, USA, 2012. ACM.

13. Stefano Marano, Vincenzo Matta, and Lang Tong. Distributed detection in the presence of byzantine attacks. *Signal Processing, IEEE Transactions on*, 57(1):16–29, 2009.
14. Achour Mostefaoui, Sergio Rajsbaum, and Michel Raynal. Conditions on input vectors for consensus solvability in asynchronous distributed systems. *Journal of the ACM (JACM)*, 50(6):922–954, 2003.
15. A. Nedic and A. Ozdaglar. Distributed subgradient methods for multi-agent optimization. *Automatic Control, IEEE Transactions on*, 54(1):48–61, Jan 2009.
16. M. Pease, R. Shostak, and L. Lamport. Reaching agreement in the presence of faults. *J. ACM*, 27(2):228–234, April 1980.
17. Boris T Poljak. *Introduction to optimization*. Optimization Software, 1987.
18. H. Robbins and D. Siegmund. A convergence theorem for non negative almost supermartingales and some applications. In T.L. Lai and D. Siegmund, editors, *Herbert Robbins Selected Papers*, pages 111–135. Springer New York, 1985.
19. Lili Su and Nitin Vaidya. Byzantine multi-agent optimization: Part I. *arXiv preprint arXiv:1506.04681*, 2015.
20. Lili Su and Nitin Vaidya. Byzantine multi-agent optimization: Part II. *CoRR*, abs/1507.01845, 2015.
21. Konstantinos I. Tsianos, Sean Lawlor, and Michael G. Rabbat. Push-sum distributed dual averaging for convex optimization. In *Decision and Control (CDC), 2012 IEEE 51st Annual Conference on*, pages 5453–5458, Dec 2012.
22. Nitin H. Vaidya. Matrix representation of iterative approximate byzantine consensus in directed graphs. *CoRR*, abs/1203.1888, 2012.
23. Nitin H Vaidya, Lewis Tseng, and Guanfeng Liang. Iterative approximate byzantine consensus in arbitrary directed graphs. In *Proceedings of the 2012 ACM symposium on Principles of distributed computing*, pages 365–374. ACM, 2012.
24. Pengfei Zhang, Jing Yang Koh, Shaowei Lin, and Ido Nevat. Distributed event detection under byzantine attack in wireless sensor networks. In *Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), 2014 IEEE Ninth International Conference on*, pages 1–6. IEEE, 2014.

Appendices

A Lemma 1

Proof. Let $x_1, x_2 \in Y$ such that $x_1 \neq x_2$. By definition of Y , there exist valid functions

$$p_1(x) = C_1 \left(\sum_{i \in \mathcal{N}} h_i(x) + \sum_{i \in \mathcal{F}} \alpha_i h_i(x) \right), \quad \text{and} \quad p_2(x) = C_2 \left(\sum_{i \in \mathcal{N}} h_i(x) + \sum_{i \in \mathcal{F}} \beta_i h_i(x) \right),$$

such that $x_1 \in \operatorname{argmin} p_1(x)$ and $x_2 \in \operatorname{argmin} p_2(x)$, respectively. Note that it is possible that $p_1(\cdot) = p_2(\cdot)$, and that $p_i(\cdot) = \tilde{p}(\cdot)$ for $i = 1$ or $i = 2$.

Given $0 \leq \alpha \leq 1$, let $x_\alpha = \alpha x_1 + (1 - \alpha)x_2$. We consider two cases:

- (i) $x_\alpha \in \operatorname{argmin} p_1(x) \cup \operatorname{argmin} p_2(x) \cup \operatorname{argmin} \tilde{p}(x)$, and
- (ii) $x_\alpha \notin \operatorname{argmin} p_1(x) \cup \operatorname{argmin} p_2(x) \cup \operatorname{argmin} \tilde{p}(x)$.

Case (i): $x_\alpha \in \operatorname{argmin} p_1(x) \cup \operatorname{argmin} p_2(x) \cup \operatorname{argmin} \tilde{p}(x)$. When $x_\alpha \in \operatorname{argmin} p_1(x) \cup \operatorname{argmin} p_2(x) \cup \operatorname{argmin} \tilde{p}(x)$, by definition of Y , we have

$$x_\alpha \in \operatorname{argmin} p_1(x) \cup \operatorname{argmin} p_2(x) \cup \operatorname{argmin} \tilde{p}(x) \subseteq Y.$$

Thus, $x_\alpha \in Y$.

Case (ii): $x_\alpha \notin \operatorname{argmin} p_1(x) \cup \operatorname{argmin} p_2(x) \cup \operatorname{argmin} \tilde{p}(x)$. By symmetry, WLOG, assume that $x_1 < x_2$. By definition of x_α and the assumption of case (ii), it holds that $x_1 < x_\alpha < x_2$. In particular, it must be that

$$x_\alpha > \max(\operatorname{argmin} p_1(x)) \text{ and } x_\alpha < \min(\operatorname{argmin} p_1(x)),$$

which imply that $p'_1(x_\alpha) > 0$ and $p'_2(x_\alpha) < 0$. There are two possibilities for $\tilde{p}'(x_\alpha)$: either $\tilde{p}'(x_\alpha) > 0$ or $\tilde{p}'(x_\alpha) < 0$. Note that $\tilde{p}'(x_\alpha) \neq 0$, since $x_\alpha \notin \operatorname{argmin} \tilde{p}(x)$.

Assume that $\tilde{p}'(x_\alpha) < 0$. Then, there exists $0 \leq \zeta \leq 1$ such that

$$\zeta p'_1(x_\alpha) + (1 - \zeta) \tilde{p}'(x_\alpha) = 0.$$

By definition of $p_1(x)$ and $\tilde{p}(x)$, we have

$$\begin{aligned} 0 &= \zeta p'_1(x_\alpha) + (1 - \zeta) \tilde{p}'(x_\alpha) \\ &= \zeta C_1 \left(\sum_{i \in \mathcal{N}} h'_i(x_\alpha) + \sum_{i \in \mathcal{F}} \alpha_i h'_i(x_\alpha) \right) + (1 - \zeta) \left(\frac{1}{|\mathcal{N}|} \sum_{i \in \mathcal{N}} h'_i(x_\alpha) \right) \\ &= \left(\zeta C_1 + (1 - \zeta) \frac{1}{|\mathcal{N}|} \right) \sum_{i \in \mathcal{N}} h'_i(x_\alpha) + \zeta C_1 \sum_{i \in \mathcal{F}} \alpha_i h'_i(x_\alpha). \end{aligned}$$

Thus, x_α is an optimum of function

$$\left(\zeta C_1 + (1 - \zeta) \frac{1}{|\mathcal{N}|} \right) \sum_{i \in \mathcal{N}} h_i(x) + \zeta C_1 \sum_{i \in \mathcal{F}} \alpha_i h_i(x). \quad (113)$$

Since $p_1(x) \in \mathcal{C}$, it holds that $C_1(|\mathcal{N}| + \sum_{i \in \mathcal{F}} \alpha_i) = 1$. Then we get

$$\begin{aligned} \left(\zeta C_1 + (1 - \zeta) \frac{1}{|\mathcal{N}|} \right) |\mathcal{N}| + \zeta C_1 \sum_{i \in \mathcal{F}} \alpha_i &= \zeta C_1 \left(|\mathcal{N}| + \sum_{i \in \mathcal{F}} \alpha_i \right) + (1 - \zeta) \frac{1}{|\mathcal{N}|} |\mathcal{N}| \\ &= \zeta 1 + (1 - \zeta) 1 = 1. \end{aligned}$$

In addition, since $\frac{1}{n} \leq C_1 \leq \frac{1}{|\mathcal{N}|}$, we get

$$\frac{1}{n} = \zeta \frac{1}{n} + (1 - \zeta) \frac{1}{n} \leq \zeta C_1 + (1 - \zeta) \frac{1}{|\mathcal{N}|} \leq \zeta \frac{1}{|\mathcal{N}|} + (1 - \zeta) \frac{1}{|\mathcal{N}|} = \frac{1}{|\mathcal{N}|}.$$

So function (113) is a valid function.

Similarly, we can show that the above result holds when $\tilde{p}'(x_\alpha) > 0$ is positive.

Therefore, set Y is convex. \square

B Lemma 2

Define an auxiliary function $r(x)$ as follows

$$r(x) \triangleq \sum_{i \in \mathcal{N}} h'_i(x) + \sum_{i \in \mathcal{F}} (h'_i(x) \mathbf{1}\{h'_i(x) > 0\}). \quad (114)$$

Proposition 4. *Function $r(x)$ is continuous and non-decreasing.*

Proof. Since $h_i(x)$ is convex for each $i \in \mathcal{V}$, it holds that $h'_i(x)$ is non-decreasing. In addition, $\mathbf{1}\{h'_i(x) > 0\}$ is also non-decreasing for each $i \in \mathcal{V}$. Thus, function $r(x)$ is non-decreasing.

For each $i \in \mathcal{V}$, since $h_i(\cdot)$ is differentiable and continuous, it follows that $h'_i(\cdot)$ is continuous. That is, $\forall \frac{\epsilon}{n} > 0$, $\exists \delta > 0$, and for each $i \in \mathcal{V}$, such that

$$|x - c| < \delta \implies |h'_i(x) - h'_i(c)| \leq \frac{\epsilon}{n}.$$

Then

$$\begin{aligned} |r(x) - r(c)| &= \left| \sum_{i \in \mathcal{N}} h'_i(x) + \sum_{i \in \mathcal{F}} (h'_i(x) \mathbf{1}\{h'_i(x) > 0\}) - \left(\sum_{i \in \mathcal{N}} h'_i(c) + \sum_{i \in \mathcal{F}} (h'_i(c) \mathbf{1}\{h'_i(c) > 0\}) \right) \right| \\ &= \left| \sum_{i \in \mathcal{N}} (h'_i(x) - h'_i(c)) + \sum_{i \in \mathcal{F}} (h'_i(x) \mathbf{1}\{h'_i(x) > 0\} - h'_i(c) \mathbf{1}\{h'_i(c) > 0\}) \right| \\ &\leq \sum_{i \in \mathcal{N}} |h'_i(x) - h'_i(c)| + \sum_{i \in \mathcal{F}} |h'_i(x) \mathbf{1}\{h'_i(x) > 0\} - h'_i(c) \mathbf{1}\{h'_i(c) > 0\}| \\ &\leq |\mathcal{N}| \frac{\epsilon}{n} + \sum_{i \in \mathcal{F}} |h'_i(x) \mathbf{1}\{h'_i(x) > 0\} - h'_i(c) \mathbf{1}\{h'_i(c) > 0\}|. \end{aligned} \quad (115)$$

When $\mathbf{1}\{h'_i(x) > 0\} = \mathbf{1}\{h'_i(c) > 0\}$, it holds that

$$\begin{aligned} |h'_i(x) \mathbf{1}\{h'_i(x) > 0\} - h'_i(c) \mathbf{1}\{h'_i(c) > 0\}| &\leq \max\{0, |h'_i(x) - h'_i(c)|\} \\ &\leq |h'_i(x) - h'_i(c)| < \frac{\epsilon}{n}. \end{aligned} \quad (116)$$

Consider the case when $\mathbf{1}\{h'_i(x) > 0\} \neq \mathbf{1}\{h'_i(c) > 0\}$. Assume $x < c$. As $h'_i(\cdot)$ is non-decreasing, we have

$$\mathbf{1}\{h'_i(x) > 0\} = 0 \neq 1 = \mathbf{1}\{h'_i(c) > 0\}.$$

Then,

$$\begin{aligned} |h'_i(x)\mathbf{1}\{h'_i(x) > 0\} - h'_i(c)\mathbf{1}\{h'_i(c) > 0\}| &= |0 - h'_i(c)| = h'_i(c) \\ &\leq h'_i(c) - h'_i(x) \quad \text{since } h'_i(x) \leq 0 \\ &= |h'_i(c) - h'_i(x)| < \frac{\epsilon}{n}. \end{aligned} \tag{117}$$

Similarly, we can show $|h'_i(x)\mathbf{1}\{h'_i(x) > 0\} - h'_i(c)\mathbf{1}\{h'_i(c) > 0\}| < \frac{\epsilon}{n}$, for the case when $x \geq c$.

By (116) and (117), we can bound (115) as

$$\begin{aligned} |r(x) - r(c)| &\leq |\mathcal{N}|\frac{\epsilon}{n} + \sum_{i \in \mathcal{F}} |h'_i(x)\mathbf{1}\{h'_i(x) > 0\} - h'_i(c)\mathbf{1}\{h'_i(c) > 0\}| \\ &\leq |\mathcal{N}|\frac{\epsilon}{n} + |\mathcal{F}|\frac{\epsilon}{n} = \epsilon. \end{aligned}$$

□

Proposition 5. *For each valid function $p(x) \in \mathcal{C}$, $\operatorname{argmin}_{x \in \mathbb{R}} p(x)$ is compact.*

Proof. Since $\operatorname{argmin}_{x \in \mathbb{R}} h_i(x)$ is compact, and $p(x)$ is a convex combination of the local functions, it follows trivially that $\operatorname{argmin}_{x \in \mathbb{R}} p(x)$ is bounded. Thus, to show $\operatorname{argmin}_{x \in \mathbb{R}} p(x)$ is compact, it remains to show that $\operatorname{argmin}_{x \in \mathbb{R}} p(x)$ is closed.

Let $\{x_t\}_{t=0}^\infty \subseteq \operatorname{argmin}_{x \in \mathbb{R}} p(x)$ be a sequence such that

$$\lim_{t \rightarrow \infty} x_t = x^*. \tag{118}$$

Recall that $h_i(\cdot)$ is continuous for each $i \in \mathcal{V}$. Then $p(x)$ is also continuous. Thus, (118) implies that

$$\lim_{t \rightarrow \infty} p(x_t) = p(x^*). \tag{119}$$

Therefore, $x^* \in \operatorname{argmin}_{x \in \mathbb{R}} p(x)$ and $\operatorname{argmin}_{x \in \mathbb{R}} p(x)$ is compact.

□

Proof of Lemma 2

Proof. By Lemma 1, we know that Y is convex. To show Y is closed, it is enough to show that Y is bounded and both $\min Y$ and $\max Y$ exist.

For small enough x , $h'_i(x) < 0$ for each $i \in \mathcal{V}$. Thus, $r(x) < 0$ for small enough x . Similarly, $r(x) > 0$ for large enough x . By Proposition 4, we know that function $r(x)$ is non-decreasing and continuous. Thus, there exists $x_0 \in \mathbb{R}$ such that

$$0 = r(x_0) = \sum_{i \in \mathcal{N}} h'_i(x_0) + \sum_{i \in \mathcal{F}} (h'_i(x_0)\mathbf{1}\{h'_i(x_0) > 0\}).$$

Let

$$p_1(x) = C_0 \left(\sum_{i \in \mathcal{N}} h_i(x) + \sum_{i \in \mathcal{F}} (h_i(x)\mathbf{1}\{h'_i(x_0) > 0\}) \right),$$

where $C_0 (|\mathcal{N}| + \sum_{i \in \mathcal{F}} \mathbf{1}\{h'_i(x_0) > 0\}) = 1$. Since

$$0 \leq \left| \sum_{i \in \mathcal{F}} \mathbf{1}\{h'_i(x_0) > 0\} \right| \leq |\mathcal{F}|,$$

it holds that $\frac{1}{n} \leq C_0 \leq \frac{1}{|\mathcal{N}|}$. Thus, $p_1(x) \in \mathcal{C}$ is a valid function.

Let $a = \min(\operatorname{argmin} p_1(x))$. By Proposition 5, $\operatorname{argmin} p_1(x)$ is compact. Thus, a is well-defined. By definition $a \in Y$. Next we show that $a = \min Y$.

Suppose, on the contrary that, there exists $\tilde{a} < a$ such that $\tilde{a} \in Y$. Since $\tilde{a} \in Y$, there exists $q(x) \in \mathcal{C}$ such that $\tilde{a} \in \operatorname{argmin} q(x)$. That is,

$$0 = q'(\tilde{a}) = C \left(\sum_{i \in \mathcal{N}} h'_i(\tilde{a}) + \sum_{i \in \mathcal{F}} \alpha_i h'_i(\tilde{a}) \right). \quad (120)$$

As $C > 0$, from (120), we have

$$\begin{aligned} 0 &= \sum_{i \in \mathcal{N}} h'_i(\tilde{a}) + \sum_{i \in \mathcal{F}} \alpha_i h'_i(\tilde{a}) \\ &\leq \sum_{i \in \mathcal{N}} h'_i(\tilde{a}) + \sum_{i \in \mathcal{F}} \alpha_i h'_i(\tilde{a}) \mathbf{1}\{h'_i(\tilde{a}) > 0\} \quad \text{since } h'_i(\tilde{a}) \mathbf{1}\{h'_i(\tilde{a}) > 0\} \geq h'_i(\tilde{a}) \\ &\leq \sum_{i \in \mathcal{N}} h'_i(\tilde{a}) + \sum_{i \in \mathcal{F}} h'_i(\tilde{a}) \mathbf{1}\{h'_i(\tilde{a}) > 0\} \quad \text{since } 0 \leq \alpha_i \leq 1 \\ &= r(\tilde{a}) \\ &\leq r(x_0) \quad \text{since } \tilde{a} < a \leq x_0 \text{ and monotonicity of } r(\cdot) \\ &= 0. \end{aligned}$$

Thus, $r(\tilde{a}) = 0 = r(x_0)$. Since $h'_i(\cdot)$ and $\mathbf{1}\{h'_i(\cdot) > 0\}$ are both non-decreasing for each $i \in \mathcal{V}$, we get

$$h'_i(\tilde{a}) = h'_i(x_0), \quad \forall i \in \mathcal{N}, \quad \text{and} \quad \mathbf{1}\{h'_i(\tilde{a}) > 0\} = \mathbf{1}\{h'_i(x_0) > 0\}, \quad \forall i \in \mathcal{F}. \quad (121)$$

We obtain

$$\begin{aligned} p'_1(\tilde{a}) &= C_1 \left(\sum_{i \in \mathcal{N}} h'_i(\tilde{a}) + \sum_{i \in \mathcal{F}} (h'_i(\tilde{a}) \mathbf{1}\{h'_i(x_0) > 0\}) \right) \\ &= C_1 \left(\sum_{i \in \mathcal{N}} h'_i(\tilde{a}) + \sum_{i \in \mathcal{F}} (h'_i(\tilde{a}) \mathbf{1}\{h'_i(\tilde{a}) > 0\}) \right) \quad \text{by (121)} \\ &= C_1 r(\tilde{a}) = 0. \end{aligned}$$

That is, $\tilde{a} \in \operatorname{argmin} p_1(x)$, contradicting the fact that $\tilde{a} < a = \min(\operatorname{argmin} p_1(x))$.

Therefore, $a = \min Y$, i.e., $\min Y$ exists. Similarly, we can show that $\max Y$ also exists.

Therefore, set Y is closed.

□

C Proof of Proposition 1

Proof. For any $t \geq 1$, we have

$$\begin{aligned}
\ell(t) &= \sum_{r=0}^{t-1} \lambda[r] b^{t-r} \\
&= \sum_{r=0}^{\lceil \frac{t}{2} \rceil} \lambda[r] b^{t-r} + \sum_{r=\lceil \frac{t}{2} \rceil+1}^{t-1} \lambda[r] b^{t-r} \\
&\leq \sum_{r=0}^{\lceil \frac{t}{2} \rceil} \lambda[0] b^{t-r} + \lambda[\lceil \frac{t}{2} \rceil] \sum_{r=\lceil \frac{t}{2} \rceil+1}^{t-1} b^{t-r} \quad \text{since } \lambda[t] \leq \lambda[t-1], \forall t \geq 1 \\
&\leq \lambda[0] \frac{b^{t-\lceil \frac{t}{2} \rceil}}{1-b} + \frac{b \lambda[\lceil \frac{t}{2} \rceil]}{1-b} \\
&\leq \lambda[0] \frac{b^{\frac{t}{2}-1}}{1-b} + \frac{b \lambda[\lceil \frac{t}{2} \rceil]}{1-b}.
\end{aligned}$$

Thus, we get

$$\limsup_{t \rightarrow \infty} \ell(t) \leq \lim_{t \rightarrow \infty} \left(\lambda[0] \frac{b^{\frac{t}{2}-1}}{1-b} + \frac{b \lambda[\lceil \frac{t}{2} \rceil]}{1-b} \right) = \lambda[0] \frac{1}{1-b} \lim_{t \rightarrow \infty} b^{\frac{t}{2}-1} + \frac{b}{1-b} \lim_{t \rightarrow \infty} \lambda[\lceil \frac{t}{2} \rceil] \stackrel{(a)}{=} 0 + 0 = 0.$$

Equality (a) follows from the fact that $0 \leq b < 1$ and the fact that $\lim_{t \rightarrow \infty} \lambda[\lceil \frac{t}{2} \rceil] = 0$. On the other hand, by definition of $\ell(t)$ we know $\ell(t) \geq 0$ for each $t \geq 1$. Thus $\liminf_{t \rightarrow \infty} \ell(t) \geq 0$.

Therefore, the limit of $\ell(t)$ exists and $\lim_{t \rightarrow \infty} \ell(t) = 0$.

□

D Proof of Lemma 12

Proof. When $t = 0$, for all $i, j \in \mathcal{N}$ we have

$$|x_i[0] - x_j[0]| \leq \max_{i \in \mathcal{N}} x_i[0] - \min_{j \in \mathcal{N}} x_j[0] = U - u.$$

Recall (94). For $t \geq 1$,

$$\mathbf{x}[t] = \mathbf{\Phi}(t-1, 0) \mathbf{x}[0] - \sum_{r=0}^{t-1} \lambda[r] \mathbf{\Phi}(t-1, r+1) \tilde{\mathbf{g}}[r],$$

Then each $x_i[t]$ can be written as

$$x_i[t] = \sum_{k=1}^{n-\phi} \mathbf{\Phi}_{ik}(t-1, 0) x_k[0] - \sum_{r=0}^{t-1} \left(\lambda[r] \sum_{k=1}^{n-\phi} \mathbf{\Phi}_{ik}(t-1, r+1) \tilde{g}_k[r] \right).$$

Thus

$$\begin{aligned}
|x_i[t] - x_j[t]| &= \left| \sum_{k=1}^{n-\phi} \Phi_{ik}(t-1, 0)x_k[0] - \sum_{r=0}^{t-1} \left(\lambda[r] \sum_{k=1}^{n-\phi} \Phi_{ik}(t-1, r+1) \tilde{g}_k[r] \right) \right. \\
&\quad \left. - \sum_{k=1}^{n-\phi} \Phi_{jk}(t-1, 0)x_k[0] + \sum_{r=0}^{t-1} \left(\lambda[r] \sum_{k=1}^{n-\phi} \Phi_{jk}(t-1, r+1) \tilde{g}_k[r] \right) \right| \\
&\leq \left| \sum_{k=1}^{n-\phi} \Phi_{ik}(t-1, 0)x_k[0] - \sum_{k=1}^{n-\phi} \Phi_{jk}(t-1, 0)x_k[0] \right| \\
&\quad + \left| \sum_{r=0}^{t-1} \left(\lambda[r] \sum_{k=1}^{n-\phi} \Phi_{jk}(t-1, r+1) \tilde{g}_k[r] \right) - \sum_{r=0}^{t-1} \left(\lambda[r] \sum_{k=1}^{n-\phi} \Phi_{ik}(t-1, r+1) \tilde{g}_k[r] \right) \right|.
\end{aligned} \tag{122}$$

We bound the two terms in (122) separately. For the first term in (122), we have

$$\begin{aligned}
\left| \sum_{k=1}^{n-\phi} \Phi_{ik}(t-1, 0)x_k[0] - \sum_{k=1}^{n-\phi} \Phi_{jk}(t-1, 0)x_k[0] \right| &= \left| \sum_{k=1}^{n-\phi} (\Phi_{ik}(t-1, 0) - \Phi_{jk}(t-1, 0)) x_k[0] \right| \\
&\leq \sum_{k=1}^{n-\phi} |\Phi_{ik}(t-1, 0) - \Phi_{jk}(t-1, 0)| |x_k[0]| \\
&\leq \sum_{k=1}^{n-\phi} \gamma^{\lceil \frac{t}{\nu} \rceil} |x_k[0]| \quad \text{by Theorem 3} \\
&\leq (n-\phi) \max\{|u|, |U|\} \gamma^{\lceil \frac{t}{\nu} \rceil}.
\end{aligned} \tag{123}$$

In addition, the second term in (122) can be bounded as follows.

$$\begin{aligned}
&\left| \sum_{r=0}^{t-1} \left(\lambda[r] \sum_{k=1}^{n-\phi} \Phi_{jk}(t-1, r+1) \tilde{g}_k[r] \right) - \sum_{r=0}^{t-1} \left(\lambda[r] \sum_{k=1}^{n-\phi} \Phi_{ik}(t-1, r+1) \tilde{g}_k[r] \right) \right| \\
&= \left| \sum_{r=0}^{t-1} \left(\lambda[r] \sum_{k=1}^{n-\phi} \Phi_{jk}(t-1, r+1) - \Phi_{ik}(t-1, r+1) \right) \tilde{g}_k[r] \right| \\
&\leq \sum_{r=0}^{t-1} \left(\lambda[r] \sum_{k=1}^{n-\phi} |\Phi_{jk}(t-1, r+1) - \Phi_{ik}(t-1, r+1)| \right) |\tilde{g}_k[r]| \\
&\leq L \sum_{r=0}^{t-1} \lambda[r] (n-\phi) \gamma^{\lceil \frac{t-1-r}{\nu} \rceil} \quad \text{by Theorem 3 and the fact that } |\tilde{g}_k[r]| \leq L
\end{aligned} \tag{124}$$

From (123) and (124), the LHS of (122) can be upper bounded by

$$|x_i[t] - x_j[t]| \leq (n-\phi) \max\{|u|, |U|\} \gamma^{\lceil \frac{t}{\nu} \rceil} + L \sum_{r=0}^{t-1} \lambda[r] (n-\phi) \gamma^{\lceil \frac{t-1-r}{\nu} \rceil}.$$

The proof is complete. \square

E Proof of Corollary 3

Proof. By Lemma 12, for each $t \geq 1$,

$$\begin{aligned} |x_i[t] - x_j[t]| &\leq (n - \phi) \max\{|u|, |U|\} \gamma^{\lceil \frac{t}{\nu} \rceil} + L \sum_{r=0}^{t-1} \lambda[r] (n - \phi) \gamma^{\lceil \frac{t-1-r}{\nu} \rceil} \\ &\leq (n - \phi) \max\{|u|, |U|\} \gamma^{\frac{t}{\nu}} + L \sum_{r=0}^{t-1} \lambda[r] (n - \phi) \gamma^{\frac{t-1-r}{\nu}}, \end{aligned}$$

and for all $i, j \in \mathcal{N}$. Taking limit sup on both sides, we get

$$\begin{aligned} \limsup_{t \rightarrow \infty} |x_i[t] - x_j[t]| &\leq (n - \phi) \max\{|u|, |U|\} \limsup_{t \rightarrow \infty} \gamma^{\frac{t}{\nu}} + L(n - \phi) \limsup_{t \rightarrow \infty} \left(\sum_{r=0}^{t-1} \lambda[r] \gamma^{\frac{t-1-r}{\nu}} \right) \\ &= 0 + L(n - \phi) \limsup_{t \rightarrow \infty} \left(\sum_{r=0}^{t-1} \lambda[r] \gamma^{\frac{t-1-r}{\nu}} \right) \\ &= 0 + 0 = 0 \quad \text{by Lemma 1,} \end{aligned}$$

proving the corollary. □

F Proof of Lemma 13

Proof. By Lemma 12, for $t \geq 1$ we have

$$M[t] - m[t] \leq (n - \phi) \max\{|u|, |U|\} \gamma^{\lceil \frac{t}{\nu} \rceil} + L \sum_{r=0}^{t-1} \lambda[r] (n - \phi) \gamma^{\lceil \frac{t-1-r}{\nu} \rceil}.$$

Thus, we get

$$\begin{aligned} \sum_{t=1}^{\infty} \lambda[t] (M[t] - m[t]) &\leq \sum_{t=1}^{\infty} \lambda[t] \left((n - \phi) \max\{|u|, |U|\} \gamma^{\lceil \frac{t}{\nu} \rceil} + L \sum_{r=0}^{t-1} \lambda[r] (n - \phi) \gamma^{\lceil \frac{t-1-r}{\nu} \rceil} \right) \\ &= (n - \phi) \max\{|u|, |U|\} \sum_{t=1}^{\infty} \lambda[t] \gamma^{\lceil \frac{t}{\nu} \rceil} + L(n - \phi) \sum_{t=1}^{\infty} \lambda[t] \sum_{r=0}^{t-1} \lambda[r] \gamma^{\lceil \frac{t-1-r}{\nu} \rceil}. \end{aligned} \tag{125}$$

Since $\lambda[t] \leq \lambda[0]$ for each $t \geq 0$, we have

$$\begin{aligned} (n - \phi) \max\{|u|, |U|\} \sum_{t=1}^{\infty} \lambda[t] \gamma^{\lceil \frac{t}{\nu} \rceil} &\leq (n - \phi) \max\{|u|, |U|\} \lambda[0] \sum_{t=1}^{\infty} \gamma^{\lceil \frac{t}{\nu} \rceil} \\ &\leq (n - \phi) \max\{|u|, |U|\} \lambda[0] \sum_{t=1}^{\infty} \gamma^{\frac{t}{\nu}} \\ &\leq (n - \phi) \max\{|u|, |U|\} \lambda[0] \frac{1}{1 - \gamma^{\frac{1}{\nu}}} < \infty. \end{aligned} \tag{126}$$

$$\begin{aligned}
L(n - \phi) \sum_{t=1}^{\infty} \lambda[t] \sum_{r=0}^{t-1} \lambda[r] \gamma^{\lceil \frac{t-1-r}{\nu} \rceil} &= L(n - \phi) \sum_{t=1}^{\infty} \sum_{r=0}^{t-1} \lambda[t] \lambda[r] \gamma^{\lceil \frac{t-1-r}{\nu} \rceil} \\
&\leq \frac{L(n - \phi)}{2} \sum_{t=1}^{\infty} \sum_{r=0}^{t-1} (\lambda^2[t] + \lambda^2[r]) \gamma^{\lceil \frac{t-1-r}{\nu} \rceil} \quad \text{since } \lambda[t] \lambda[r] \leq \frac{\lambda^2[t] + \lambda^2[r]}{2} \\
&= \frac{L(n - \phi)}{2} \sum_{t=1}^{\infty} \lambda^2[t] \sum_{r=0}^{t-1} \gamma^{\lceil \frac{t-1-r}{\nu} \rceil} + \frac{L(n - \phi)}{2} \sum_{t=1}^{\infty} \sum_{r=0}^{t-1} \lambda^2[r] \gamma^{\lceil \frac{t-1-r}{\nu} \rceil}
\end{aligned} \tag{127}$$

The first term on the RHS of (127) can be bounded as

$$\begin{aligned}
\frac{L(n - \phi)}{2} \sum_{t=1}^{\infty} \lambda^2[t] \sum_{r=0}^{t-1} \gamma^{\lceil \frac{t-1-r}{\nu} \rceil} &\leq \frac{L(n - \phi)}{2} \sum_{t=1}^{\infty} \lambda^2[t] \sum_{r=0}^{t-1} \gamma^{\frac{t-1-r}{\nu}} \\
&\leq \frac{L(n - \phi)}{2} \sum_{t=1}^{\infty} \lambda^2[t] \frac{1}{1 - \gamma^{\frac{1}{\nu}}} \\
&= \frac{L(n - \phi)}{2 \left(1 - \gamma^{\frac{1}{\nu}}\right)} \sum_{t=1}^{\infty} \lambda^2[t] \\
&< \infty \quad \text{since } \sum_{t=1}^{\infty} \lambda^2[t] < \infty.
\end{aligned} \tag{128}$$

For the second term on the RHS of (127), for any fixed T , we get

$$\begin{aligned}
\frac{L(n - \phi)}{2} \sum_{t=1}^T \sum_{r=0}^{t-1} \lambda^2[r] \gamma^{\lceil \frac{t-1-r}{\nu} \rceil} &\leq \frac{L(n - \phi)}{2} \sum_{t=1}^T \sum_{r=0}^{t-1} \lambda^2[r] \gamma^{\frac{t-1-r}{\nu}} \\
&= \frac{L(n - \phi)}{2} \sum_{r=0}^{T-1} \lambda^2[r] \sum_{t=0}^{T-1-r} \gamma^{\frac{t}{\nu}} \\
&\leq \frac{L(n - \phi)}{2(1 - \gamma^{\frac{1}{\nu}})} \sum_{r=0}^{T-1} \lambda^2[r].
\end{aligned}$$

Thus, we get

$$\frac{L(n - \phi)}{2} \sum_{t=0}^{\infty} \sum_{r=0}^{t-1} \lambda^2[r] \gamma^{\lceil \frac{t-1-r}{\nu} \rceil} \leq \frac{L(n - \phi)}{2(1 - \gamma^{\frac{1}{\nu}})} \sum_{r=0}^{\infty} \lambda^2[r] < \infty. \tag{129}$$

By (126), (128) and (129), we get

$$\sum_{t=1}^{\infty} \lambda[t] (M[t] - m[t]) < \infty.$$

In addition,

$$\begin{aligned}
\sum_{t=0}^{\infty} \lambda[t] (M[t] - m[t]) &= \lambda[0] (M[0] - m[0]) + \sum_{t=1}^{\infty} \lambda[t] (M[t] - m[t]) \\
&= \lambda[0](U - u) + \sum_{t=1}^{\infty} \lambda[t] (M[t] - m[t]) < \infty,
\end{aligned}$$

proving the lemma. □

G Proof of Lemma 11

Define an auxiliary function $\tilde{r}(x)$ as follows. For each $x \in \mathbb{R}$, let $h'_{i_1(x)}(x), \dots, h'_{i_{|\mathcal{N}|}(x)}(x)$ be a non-decreasing order of $h'_j(x)$, for $j \in \mathcal{N}$. Define $\tilde{r}(x)$ as follows,

$$\tilde{r}(x) = \left(1 - \frac{|\mathcal{N}| - f - 1}{2(|\mathcal{N}| - f)}\right) h'_{i_1(x)}(x) + \frac{1}{2(|\mathcal{N}| - f)} \sum_{j=2}^{|\mathcal{N}|-f} h'_{i_j(x)}(x). \quad (130)$$

Intuitively speaking, $\tilde{r}(x)$ is the largest gradient value among all valid functions in \tilde{C} at point x .

Proposition 6. *Function $\tilde{r}(\cdot)$ is continuous and non-decreasing.*

Proof. Let $x \leq y \in \mathbb{R}$.

$$\begin{aligned} \tilde{r}(y) - \tilde{r}(x) &= \left(1 - \frac{|\mathcal{N}| - f - 1}{2(|\mathcal{N}| - f)}\right) h'_{i_1(y)}(y) + \frac{1}{2(|\mathcal{N}| - f)} \sum_{j=2}^{|\mathcal{N}|-f} h'_{i_j(y)}(y) \\ &\quad - \left(1 - \frac{|\mathcal{N}| - f - 1}{2(|\mathcal{N}| - f)}\right) h'_{i_1(x)}(x) - \frac{1}{2(|\mathcal{N}| - f)} \sum_{j=2}^{|\mathcal{N}|-f} h'_{i_j(x)}(x) \\ &\geq \left(1 - \frac{|\mathcal{N}| - f - 1}{2(|\mathcal{N}| - f)}\right) h'_{i_1(x)}(y) + \frac{1}{2(|\mathcal{N}| - f)} \sum_{j=2}^{|\mathcal{N}|-f} h'_{i_j(x)}(y) \\ &\quad - \left(1 - \frac{|\mathcal{N}| - f - 1}{2(|\mathcal{N}| - f)}\right) h'_{i_1(x)}(x) - \frac{1}{2(|\mathcal{N}| - f)} \sum_{j=2}^{|\mathcal{N}|-f} h'_{i_j(x)}(x) \\ &= \left(1 - \frac{|\mathcal{N}| - f - 1}{2(|\mathcal{N}| - f)}\right) \left(h'_{i_1(x)}(y) - h'_{i_1(x)}(x)\right) + \frac{1}{2(|\mathcal{N}| - f)} \sum_{j=2}^{|\mathcal{N}|-f} \left(h'_{i_j(x)}(y) - h'_{i_j(x)}(x)\right) \\ &\leq 0 + 0 \quad \text{since } x \leq y \text{ and } h'_i(\cdot) \text{ is non-decreasing} \end{aligned}$$

Thus, function $\tilde{r}(\cdot)$ is non-decreasing.

Next we show that function $\tilde{r}(\cdot)$ is continuous.

For each $i \in \mathcal{V}$, since $h_i(\cdot)$ is differentiable, it follows that $h'_i(\cdot)$ is continuous. That is, $\forall \epsilon > 0$, $\exists \delta > 0$, and for each $i \in \mathcal{V}$, such that

$$|x - c| < \delta \implies |h'_i(x) - h'_i(c)| \leq \epsilon.$$

Assume $c \leq x < c + \delta$. Then

$$\begin{aligned}
|\tilde{r}(x) - \tilde{r}(c)| &= \tilde{r}(x) - \tilde{r}(c) \quad \text{by monotonicity of } \tilde{r}(\cdot) \\
&= \left(1 - \frac{|\mathcal{N}| - f - 1}{2(|\mathcal{N}| - f)}\right) h'_{i_1(x)}(x) + \frac{1}{2(|\mathcal{N}| - f)} \sum_{j=2}^{|\mathcal{N}|-f} h'_{i_j(x)}(x) \\
&\quad - \left(1 - \frac{|\mathcal{N}| - f - 1}{2(|\mathcal{N}| - f)}\right) h'_{i_1(c)}(c) - \frac{1}{2(|\mathcal{N}| - f)} \sum_{j=2}^{|\mathcal{N}|-f} h'_{i_j(c)}(c) \\
&\leq \left(1 - \frac{|\mathcal{N}| - f - 1}{2(|\mathcal{N}| - f)}\right) h'_{i_1(x)}(x) + \frac{1}{2(|\mathcal{N}| - f)} \sum_{j=2}^{|\mathcal{N}|-f} h'_{i_j(x)}(x) \\
&\quad - \left(1 - \frac{|\mathcal{N}| - f - 1}{2(|\mathcal{N}| - f)}\right) h'_{i_1(x)}(c) - \frac{1}{2(|\mathcal{N}| - f)} \sum_{j=2}^{|\mathcal{N}|-f} h'_{i_j(x)}(c) \\
&\leq \left(1 - \frac{|\mathcal{N}| - f - 1}{2(|\mathcal{N}| - f)}\right) (h'_{i_1(x)}(x) - h'_{i_1(x)}(c)) + \frac{1}{2(|\mathcal{N}| - f)} \sum_{j=2}^{|\mathcal{N}|-f} (h'_{i_j(x)}(x) - h'_{i_j(x)}(c)) \\
&< \left(1 - \frac{|\mathcal{N}| - f - 1}{2(|\mathcal{N}| - f)}\right) \epsilon + \frac{1}{2(|\mathcal{N}| - f)} \sum_{j=2}^{|\mathcal{N}|-f} \epsilon = \epsilon. \tag{131}
\end{aligned}$$

Similarly, we can show that when $c - \delta < x \leq c$, $|\tilde{r}(x) - \tilde{r}(c)| < \epsilon$.

Thus, function $\tilde{r}(\cdot)$ is continuous.

The proof is complete. □

Proof of Lemma 11

Proof. By Lemma 10, we know that \tilde{Y} is convex. To show Y is closed, it is enough to show that \tilde{Y} is bounded and both $\min \tilde{Y}$ and $\max \tilde{Y}$ exist.

By Proposition 6, we know that function $\tilde{r}(x)$ is non-decreasing and continuous. Thus, there exists $x_0 \in \mathbb{R}$ such that

$$0 = \tilde{r}(x_0) = \left(1 - \frac{|\mathcal{N}| - f - 1}{2(|\mathcal{N}| - f)}\right) h'_{i_1(x_0)}(x_0) + \frac{1}{2(|\mathcal{N}| - f)} \sum_{j=2}^{|\mathcal{N}|-f} h'_{i_j(x_0)}(x_0).$$

Let

$$q(x) = \left(1 - \frac{|\mathcal{N}| - f - 1}{2(|\mathcal{N}| - f)}\right) h_{i_1(x_0)}(x) + \frac{1}{2(|\mathcal{N}| - f)} \sum_{j=2}^{|\mathcal{N}|-f} h_{i_j(x_0)}(x). \tag{132}$$

By construction, $q(x) \in \mathcal{C}$ is a valid function. Note that due to the possibility of existence of ties in top $|\mathcal{N}| - f$ rankings of the order $h'_{i_1(x)}(x), \dots, h'_{i_{|\mathcal{N}|}(x)}(x)$, for a given x_0 , there may be multiple orders over $h'_i(x_0), \forall i \in \mathcal{N}$ of the top $|\mathcal{N}| - f$ elements. Let \mathcal{O} be the collection of all such orders.

Note that there is an one-to-one correspondence of an order and a valid function defined in (132). Let

$$a = \min_{o \in \mathcal{O}} \min (\operatorname{argmin} q_o(x)),$$

which is well-defined since $\operatorname{argmin} q_o(x)$ is compact, and $|\mathcal{O}|$ is finite.

By definition $a \in \tilde{Y}$. Next we show that $a = \min \tilde{Y}$.

Suppose, on the contrary that, there exists $\tilde{a} < a$ such that $\tilde{a} \in \tilde{Y}$. Since $\tilde{a} \in \tilde{Y}$, there exists $\tilde{q}(x) = \sum_{i \in \mathcal{N}} \alpha_i h_i(x) \in \tilde{\mathcal{C}}$ such that $\tilde{a} \in \operatorname{argmin} \tilde{q}(x)$. That is,

$$\tilde{q}'(\tilde{a}) = 0. \quad (133)$$

We have

$$\begin{aligned} 0 = q'(\tilde{a}) &= \sum_{i \in \mathcal{N}} \alpha_i h'_i(\tilde{a}) \\ &\leq \left(1 - \frac{|\mathcal{N}| - f - 1}{2(|\mathcal{N}| - f)}\right) h'_{i_1(\tilde{a})}(\tilde{a}) + \frac{1}{2(|\mathcal{N}| - f)} \sum_{j=2}^{|\mathcal{N}| - f} h'_{i_j(\tilde{a})}(\tilde{a}) \\ &= \tilde{r}(\tilde{a}) \leq \tilde{r}(x_0) = 0 \quad \text{by monotonicity of } \tilde{r}(\cdot) \end{aligned}$$

Thus, $\tilde{r}(\tilde{a}) = 0 = \tilde{r}(x_0)$. In addition, we have

$$\begin{aligned} 0 &= \left(1 - \frac{|\mathcal{N}| - f - 1}{2(|\mathcal{N}| - f)}\right) h'_{i_1(\tilde{a})}(\tilde{a}) + \frac{1}{2(|\mathcal{N}| - f)} \sum_{j=2}^{|\mathcal{N}| - f} h'_{i_j(\tilde{a})}(\tilde{a}) \\ &\leq \left(1 - \frac{|\mathcal{N}| - f - 1}{2(|\mathcal{N}| - f)}\right) h'_{i_1(\tilde{a})}(x_0) + \frac{1}{2(|\mathcal{N}| - f)} \sum_{j=2}^{|\mathcal{N}| - f} h'_{i_j(\tilde{a})}(x_0) \quad \text{by monotonicity of } h'_i(\cdot) \\ &\leq \left(1 - \frac{|\mathcal{N}| - f - 1}{2(|\mathcal{N}| - f)}\right) h'_{i_1(x_0)}(x_0) + \frac{1}{2(|\mathcal{N}| - f)} \sum_{j=2}^{|\mathcal{N}| - f} h'_{i_j(x_0)}(x_0), \end{aligned}$$

which implies that $i_1(\tilde{a}), \dots, i_{|\mathcal{N}| - f}(\tilde{a})$ is an order in \mathcal{O} . Thus, it can be seen that $\tilde{a} \geq a = \min_{o \in \mathcal{O}} \min (\operatorname{argmin} q_o(x))$, contradicting the assumption that $\tilde{a} < a$.

Therefore, $a = \min \tilde{Y}$, i.e., $\min \tilde{Y}$ exists. Similarly, we can show that $\max \tilde{Y}$ also exists.

Therefore, set \tilde{Y} is closed.

□